

(19)



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 0 321 362 B1**

(12)

**EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention  
of the grant of the patent:  
**25.09.1996 Bulletin 1996/39**

(51) Int Cl.<sup>6</sup>: **C12N 15/12, C12P 21/02,  
C12Q 1/68, C07K 14/00**

(21) Application number: **88403229.3**

(22) Date of filing: **16.12.1988**

(54) **Retinoic acid receptor and derivatives thereof, DNA encoding either substance and use of the proteins and of the DNA**

Retinoesäurerezeptor und Derivate davon, beide Substanzen codierende DNS und die Verwendung der Proteine und der DNS

Récepteur de l'acide rétinolique et ses dérivées, DNA codant pour ces deux substances et l'usage des protéines et des DNA

(84) Designated Contracting States:  
**AT BE CH DE ES FR GB GR IT LI LU NL SE**

(30) Priority: **16.12.1987 US 133687  
17.12.1987 US 134130  
20.06.1988 US 209009  
30.11.1988 US 278136**

(43) Date of publication of application:  
**21.06.1989 Bulletin 1989/25**

(73) Proprietor: **INSTITUT PASTEUR  
F-75715 Paris Cedex 15 (FR)**

(72) Inventors:  
• **Tioilals, Pierre**  
F-75013 Paris (FR)  
• **Dejean, Anne**  
F-75014 Paris (FR)  
• **Blaudin de The, Hugues**  
F-75003 Paris (FR)  
• **Marchio, Agnès**  
F-75011 Paris (FR)

(74) Representative: **Gutmann, Ernest et al**  
**Ernest Gutmann - Yves Plasseraud S.A.**  
**3, rue Chauveau-Lagarde**  
**75008 Paris (FR)**

(56) References cited:  
**EP-A- 0 244 221**

- **NATURE**, vol. 333, 16th June 1988, pages 669-672; D. BENBROOK et al.: "A new retinoic acid receptor identified from a hepatocellular carcinoma"
- **NATURE**, vol. 332, 28th April 1988, pages 850-853; N. BRAND et al.: "Identification of a second human retinoic acid receptor"
- **NATURE**, vol. 322, 3rd July 1986, pages 70-72; A. DEJEAN et al.: "Hepatitis B virus DNA intergration in a sequence homologous to v-erb-A and steroid receptor genes in a hepatocellular carcinoma"
- **NATURE**, vol. 330, 3rd December 1987, pages 444-450; M. PETKOVICH et al.: "A human retinoic acid receptor which belongs to the family of nuclear receptors"
- **SCIENCE**, vol. 240, 13th May 1988, pages 889-895; R.M. EVANS: "The steroid and thyroid hormone receptor superfamily"
- **NATURE**, vol. 330, 17th December 1987, pp. 667-670; H. DE THE.

**EP 0 321 362 B1**

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

**Description****BACKGROUND OF THE INVENTION**

5 This invention relates to nucleotide sequences, polypeptides encoded by the nucleotide sequences, and to their use in diagnostic and pharmaceutical applications.

Primary hepatocellular carcinoma (HCC) represents the most common cancer, especially in young men, in many parts of the world (as in China and in much of Asia and Africa) (reviewed in Tiollais et al., 1985). Its etiology was investigated mostly by epidemiological studies, which revealed that, beyond some minor potential agents such as aflatoxin and sex steroids, hormones, Hepatitis B virus (HBV) chronic infection could account for a large fraction of liver cancers (Beasley and Hwang, 1984).

10 HBV DNA has been found to be integrated in the genome of most cases of HCCs studied (Edman et al., 1980; Brechot et al., 1980; Chakraborty et al., 1980; Chen et al., 1982). Nonetheless the role of those sequences in liver oncogenesis remains unclear.

15 A single HBV integration in a HCC sample in a short liver cell sequence has been reported recently. The sequence was found to be homologous to steroid receptor genes and to the cellular proto-oncogene *c-erbA* (Dejean et al., 1986).

Ligand-dependent transcriptional activators, such as steroid or thyroid hormone receptors, have recently been cloned allowing rapid progress in the understanding of their mechanism of action. Nevertheless, there exists a need in the art for the identification of transcripts that may encode for activational elements, such as nuclear surface receptors, that may play a role in hepatocellular carcinoma. Such findings would aid in identifying corresponding transcripts in susceptible individuals. In addition, identification of transcripts could aid in elucidating the mechanisms by which HCC occurs.

20 Retinoids, a class of compounds including retinol (vitamin A), retinoic acid (RA), and a series of natural and synthetic derivatives, exhibit striking effects on cell proliferation, differentiation, and pattern formation during development (Strickland and Mahdavi, 1978; Breitman et al., 1980; Roberts and Sporn, 1984; Thaller and Eichele, 1987). Until recently, the molecular mechanism by which these compounds exert such potent effects was unknown, although retinoids were thought to modify their target cells through a specific receptor.

25 Except for the role of retinoids in vision, their mechanism of action is not well understood at the molecular level. Several possible mechanisms have been suggested. One hypothesis proposes that retinoids are needed to serve as the lipid portion of glycolipid intermediates involved in certain, specific glycosylation reactions. Another mechanism, which may account for the various effects of retinoids on target cells, is that they alter genomic expression in such cells. It has been suggested that retinoids may act in a manner analogous to that of the steroid hormones and that the intracellular binding proteins (cellular retinal-binding and retinoic acid-binding protein) play a critical part in facilitating the interaction of retinoids with binding sites in the cell nucleus.

30 For example, the observation that the RA-induced differentiation of murine F9 embryonal carcinoma cells is accompanied by the activation of specific genes has led to the proposal that RA, like the steroid and thyroid hormones, could exert its transcriptional control by binding to a nuclear receptor (Roberts and Spron, 1984). However, the biochemical characterization of this receptor had been hampered by high affinity RA-binding sites corresponding to the cellular retinoic acid binding protein (CRABP), which is thought to be a cytoplasmic shuttle for RA (Chytil and Ong, 1984).

35 In any event, retinoids are currently of interest in dermatology. The search for new retinoids has identified a number of compounds with a greatly increased therapeutic index as compared with naturally occurring retinoids. Extensive clinical testing of two of these retinoids, 13-*cis*-retinoic acid and the aromatic analog etretinate, has lead to their clinical use in dermatology. In addition, several lines of evidence suggest that important relations exist between retinoids and cancer. A number of major diseases, in addition to cancer, are characterized by excessive proliferation of cells, often with excessive accumulation of extracellular matrix material. These diseases include rheumatoid arthritis, psoriasis, idiopathic pulmonary fibrosis, scleroderma, and cirrhosis of the liver, as well as the disease process atherosclerosis. The possibility exists that retinoids, which can influence cell differentiation and proliferation, may be of therapeutic value in some of the proliferative diseases. There exists a need in the art for reagents and methods for carrying out studies of receptor expression and effector function to determine whether candidate drugs are agonists or antagonists of retinoid activity in biological systems.

40 There also exists a need in the art for identification of retinoic acid receptors and for sources of retinoic acid receptors in highly purified form. The availability of the purified receptor would make it possible to assay fluids for agonists and antagonists of the receptor.

**SUMMARY OF THE INVENTION**

55 This invention aids in fulfilling these needs in the art. More particularly, this invention provides a cloned DNA

sequence encoding for a polypeptide of a newly identified cellular gene, which has been named hap. The DNA sequence has the formula shown in figures 2a, 2b, 2c, 2d successively (and collectively designated as figure 2). More particularly, the sequence comprises a sequence encoding for a polypeptide named hap or a fragment thereof having the following formula :

5

```

    ATG TTT GAC TGT ATG GAT GTT CTG TCA GTG AGT CCT GGG CAA
    ATC CTG GAT TTC TAC ACT GCG AGT CCG TCT TCC TGC ATG CTC
10  CAG GAG AAA GCT CTC AAA GCA TGC TTC AGT GGA TTG ACC CAA
    ACC GAA TGG CAG CAT CGG CAC ACT GCT CAA TCA ATT GAA ACA
    CAG AGC ACC AGC TCT GAG GAA CTC GTC CCA AGC CCC CCA TCT
15  CCA CTT CCT CCC CCT CGA GTG TAC AAA CCC TGC TTC GTC TGC
    CAG GAC AAA TCA TCA GGG TAC CAC TAT GGG GTC AGC GCC TGT
    GAG GGA TGT AAG GGC TTT TTC CGC AGA AGT ATT CAG AAG AAT
20  ATG ATT TAC ACT TGT CAC CGA GAT AAG AAC TGT GTT ATT AAT
    AAA GTC ACC AGG AAT CGA TGC CAA TAC TGT CGA CTC CAG AAG
    TGC TTT GAA GTG GGA ATG TCC AAA GAA TCT GTC AGG AAT GAC
    AGG AAC AAG AAA AAG AAG GAG ACT TCG AAG CAA GAA TGC ACA
25  GAG AGC TAT GAA ATG ACA GCT GAG TTG GAC GAT CTC ACA GAG
    AAG ATC CGA AAA GCT CAC CAG GAA ACT TTC CCT TCA CTC TGC
    CAG CTG GGT AAA TAC ACC ACG AAT TCC AGT GCT GAC CAT CGA
30  GTC CGA CTG GAC CTG GGC CTC TGG GAC AAA TTC AGT GAA CTG
    GCC ACC AAG TGC ATT ATT AAG ATC GTG GAG TTT GCT AAA CGT
    CTG CCT GGT TTC ACT GGC TTG ACC ATC GCA GAC CAA ATT ACC
35  CTG CTG AAG GCC GCC TGC CTG GAC ATC CTG ATT CTT AGA ATT
    TGC ACC AGG TAT ACC CCA GAA CAA GAC ACC ATG ACT TTC TCA
    GAC GGC CTT ACC CTA AAT CGA ACT CAG ATG CAC AAT GCT GGA
40  TTT GGT CCT CTG ACT GAC CTT GTG TTC ACC TTT GCC AAC CAG
    CTC CTG CCT TTG GAA ATG GAT GAC ACA GAA ACA GGC CTT CTC
    AGT GCC ATC TGC TTA ATC TGT GGA GAC CGC CAG GAC CTT GAG
    GAA CCG ACA AAA GTA GAT AAG CTA CAA GAA CCA TTG CTG GAA
45  GCA CTA AAA ATT TAT ATC AGA AAA AGA CGA CCC AGC AAG CCT
    CAC ATG TTT CCA AAG ATC TTA ATG AAA ATC ACA GAT CTC CGT
    AGC ATC AGT GCT AAA GGT GCA GAG CGT GTA ATT ACC TTG AAA
50  ATG GAA ATT CCT GGA TCA ATG CCA CCT CTC ATT CAA GAA ATG
    ATG GAG AAT TCT GAA GGA CAT GAA CCC TTG ACC CCA AGT TCA
    AGT GGG AAC ACA GCA GAG CAC AGT CCT AGC ATC TCA CCC AGC
55  TCA GTG GAA AAC AGT GGG GTC AGT CAG TCA CCA CTC GTG CAA
    TAA,

```

The invention also covers variants and fragments of the DNA sequence. The DNA sequence is in a purified form.

This invention also provides a probe consisting of a radionucléotide bonded to the DNA sequence of the invention.

In addition, this invention provides a hybrid duplex molecule consisting essentially of the DNA sequence of the invention hydrogen bonded to a nucleotide sequence of complementary base sequence, such as DNA or RNA.

5 Further, this invention provides a polypeptide comprising an amino acid sequence of hap protein, wherein the polypeptide contains the amino acid sequence shown in Figure 2. More particularly, the amino acid sequence consists of the following sequence :

10 Met Phe Asp Cys Met Asp Val Leu Ser Val Ser Pro Gly Gln  
 Ile Leu Asp Phe Tyr Thr Ala Ser Pro Ser Ser Cys Met Leu  
 Gln Glu Lys Ala Leu Lys Ala Cys Phe Ser Gly Leu Thr Gln  
 Thr Glu Trp Gln His Arg His Thr Ala Gln Ser Ile Glu Thr  
 15 Gln Ser Thr Ser Ser Glu Glu Leu Val Pro Ser Pro Pro Ser  
 Pro Leu Pro Pro Pro Arg Val Tyr Lys Pro Cys Phe Val Cys  
 Gln Asp Lys Ser Ser Gly Tyr His Tyr Gly Val Ser Ala Cys  
 20 Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Lys Asn  
 Met Ile Tyr Thr Cys His Arg Asp Lys Asn Cys Val Ile Asn  
 Lys Val Thr Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys  
 Cys Phe Glu Val Gly Met Ser Lys Glu Ser Val Arg Asn Asp  
 25 Arg Asn Lys Lys Lys Lys Glu Thr Ser Lys Gln Glu Cys Thr  
 Glu Ser Tyr Glu Met Thr Ala Glu Leu Asp Asp Leu Thr Glu  
 Lys Ile Arg Lys Ala His Gln Glu Thr Phe Pro Ser Leu Cys  
 30 Gln Leu Gly Lys Tyr Thr Thr Asn Ser Ser Ala Asp His Arg  
 Val Arg Leu Asp Leu Gly Leu Trp Asp Lys Phe Ser Glu Leu  
 Ala Thr Lys Cys Ile Ile Lys Ile Val Glu Phe Ala Lys Arg  
 35 Leu Pro Gly Phe Thr Gly Leu Thr Ile Ala Asp Gln Ile Thr  
 Leu Leu Lys Ala Ala Cys Leu Asp Ile Leu Ile Leu Arg Ile

40

45

50

55

Cys Thr Arg Tyr Thr Pro Glu Gln Asp Thr Met Thr Phe Ser  
 Asp Gly Leu Thr Leu Asn Arg Thr Gln Met His Asn Ala Gly  
 5 Phe Gly Pro Leu Thr Asp Leu Val Phe Thr Phe Ala Asn Gln  
 Leu Leu Pro Leu Glu Met Asp Asp Thr Glu Thr Gly Leu Leu  
 Ser Ala Ile Cys Leu Ile Cys Gly Asp Arg Gln Asp Leu Glu  
 10 Glu Pro Thr Lys Val Asp Lys Leu Gln Glu Pro Leu Leu Glu  
 Ala Leu Lys Ile Tyr Ile Arg Lys Arg Arg Pro Ser Lys Pro  
 His Met Phe Pro Lys Ile Leu Met Lys Ile Thr Asp Leu Arg  
 15 Ser Ile Ser Ala Lys Gly Ala Glu Arg Val Ile Thr Leu Lys  
 Met Glu Ile Pro Gly Ser Met Pro Pro Leu Ile Gln Glu Met  
 Met Glu Asn Ser Glu Gly His Glu Pro Leu Thr Pro Ser Ser  
 Ser Gly Asn Thr Ala Glu His Ser Pro Ser Ile Ser Pro Ser  
 20 Ser Val Glu Asn Ser Gly Val Ser Gln Ser Pro Leu Val Gln

The invention also covers serotypic variants of the polypeptide and fragments of the polypeptide. The polypeptide is free from human serum proteins, virus, viral proteins, human tissue, and human tissue components. Preferably, the polypeptide is free from human, blood-derived protein.

The hap protein (hap for hepatoma) exhibits strong homology with the human retinoic acid receptor (RAR) de The, H., Marchio, A., Tiollais, P. & Dejean, A. *Nature* 330, 667-670 (1987), Petkovich, M., Brand, N.J., Krust, A. & Chambon, P. *Nature* 330, 444-450 (1987), Giguere, V., Ong, E.S., Segui, P. & Evans, R.M. *Nature* 330, 624-629 (1987). To test the possibility that the hap protein might also be a retinoid receptor, a chimaeric receptor was created by replacing the putative DNA binding domain of hap with that of the human oestrogen receptor (ER). The resulting hap-ER chimaera was then tested for its ability to trans-activate an oestrogen-responsive reporter gene (vit-tk-CAT) in the presence of possible receptor ligands. It was discovered that retinoic acid (RA) at physiological concentrations is effective in inducing the expression of this reporter gene by the hap-ER chimaeric receptor. See *Nature*, 332:850-853 (1988). This demonstrates the existence of two human retinoic acid receptors designated RAR- $\alpha$  and RAR- $\beta$ .

More particularly, it has been discovered that the hap protein is a second retinoic acid receptor. Thus, the expression "hap protein" is used interchangeably herein with the abbreviation "RAR- $\beta$ " for the second human retinoic acid receptor.

Also, this invention provides a process for selecting a nucleotide sequence coding for hap protein or a portion thereof from a group of nucleotide sequences comprising the step of determining which of the nucleotide sequences hybridizes to a DNA sequence of the invention. The nucleotide sequence can be a DNA sequence or an RNA sequence. The process can include the step of detecting a label on the nucleotide sequence.

Still further, this invention provides a recombinant vector comprising lambda-NH1149 having an EcoRI restriction endonuclease site into which has been inserted the DNA sequence of the invention. The invention also provides plasmid pCOD20, which comprises the DNA sequence of the invention.

This invention provides an E. coli bacterial culture in a purified form, wherein the culture comprises E. coli cells containing DNA, wherein a portion of the DNA comprises the DNA sequence of the invention. Preferably, the E. coli is strain TG-1.

In addition, this invention provides a method of using the purified retinoic acid receptor of the invention for assaying a medium, such as a fluid, for the presence of an agonist or antagonist of the receptor. In general, the method comprises providing a known concentration of a proteaceous receptor of the invention, incubating the receptor with a ligand of the receptor and a suspected agonist or antagonist under conditions sufficient to form a ligand-receptor complex, and assaying for ligand-receptor complex or for free ligand or for non-complex receptor. The assay can be conveniently carried out using labelled reagents as more fully described hereinafter, and conventional techniques based on nucleic acid hybridization, immunochemistry, and chromatography, such as TLC, HPLC, and affinity chromatography.

In another method of the invention, a medium is assayed for stimulation of transcription of the RAR- $\beta$  gene or translation of the gene by an agonist or antagonist. For example,  $\beta$ -receptor binding retinoids can be screened in this manner.

BRIEF DESCRIPTION OF THE DRAWINGS

This invention will be described in greater detail with reference to the drawings in which

- 5 Fig. 1 is a restriction map of human liver hap cDNA;
- Fig. 2 (figs 2a, 2b, 2c, 2d) is the nucleotide sequence of human liver hap cDNA and a predicted amino acid sequence of human liver hap cDNA;
- 10 Fig. 3 depicts the distribution of hap mRNA in different tissues as determined by Northern blot analysis;
- Fig. 4 depicts the distribution of hap mRNA in HCC and HCC derived cell lines as determined by Northern blot analysis;
- 15 Fig. 5 is a fluorograph of hap polypeptide synthesized in vitro and isolated on SDS-polyacrylamide gel;
- Fig. 6 shows the alignment of hap translated amino acid sequence with several known sequences for thyroid and steroid hormone receptors;
- 20 Fig. 7 is a schematic alignment of similar regions identified as A/B, C, D, and E of the amino acid sequences of Fig. 6;
- Fig. 8 depicts hap related genes in vertebrates (A) and in humans (B and C) as determined by Southern blot analysis;
- 25 Fig. 9 shows the tissue distribution of RAR  $\alpha$  and  $\beta$  transcripts;
- Fig. 10 shows the dose- and time-response of RAR  $\alpha$  and  $\beta$  transcripts after retinoic acid treatment of PLC/PRF/5 cells;
- 30 Fig. 11 shows the effect of RNA and protein synthesis inhibitors on the levels of RAR  $\alpha$  and  $\beta$  mRNAs;
- Fig. 12 reports the results of nuclear run-on analysis of RAR  $\beta$  gene transcription after RA treatment; and
- Fig. 13 reports the results of nuclear run-on analysis of RAR  $\beta$  transcription in two hepatoma cell-lines;
- 35 Fig. 14 shows the resulting kinetic analysis of RAR mRNA degradation;
- Fig. 15 depicts a nucleotide sequence analysis extending a  $\lambda$  13 RAR- $\beta$  by 72 bp; and
- 40 Fig. 16 is a complete restriction map of a cloned HindIII-BamHI genomic DNA insert containing the nucleotide sequence of Fig. 15.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

45 A. IDENTIFICATION OF A PROTEIN, NAMED hap PROTEIN, HAVING DNA-BINDING AND LIGAND-BINDING DOMAINS, AND IDENTIFICATION OF THE DNA SEQUENCE ENCODING hap PROTEIN

As previously noted, ligand-dependent transcriptional activators, such as steroid or thyroid hormone receptors, have recently been cloned. The primary structure and expression of a new gene, hap, closely related to steroid or thyroid hormone receptor genes have now been discovered. The hap product exhibits two regions highly homologous to the conserved DNA- and hormone-binding domains of previously cloned receptors.

50 More particularly, the cloning of a cDNA corresponding to a novel steroid/thyroid hormone receptor-related gene has been achieved. The cDNA was recovered from a human liver cDNA library using a labelled cellular DNA fragment previously isolated from a liver tumor. The fragment contained a 147 bp putative exon in which HBV inserted. The sequence of this cellular gene, which is referred to herein as hap for hepatoma, reveals various structural features characteristic of c-erbA/steroid receptors (Dejean et al., 1986). The receptor-related protein is likely to be a novel member of the superfamily of transcriptional regulatory proteins that includes the thyroid and steroid hormone receptors.

55 It has been discovered that the hap gene is transcribed at low level in most human tissues, but the gene is over-

expressed in prostate and kidney. Moreover, six out of seven hepatoma and hepatoma-derived cell lines express a small hap transcript, which is undetectable in normal adult and fetal livers, but is present in all non-hepatic tissues tested. Altered expression of hap may be involved in liver oncogenesis.

These findings, as well as other discoveries relating to this invention, will now be described in detail.

#### A.1 Cloning and Sequencing of a Hap cDNA

A human liver cDNA library was screened using a nick-translated 350 bp EcoRI genomic fragment (MNT probe) previously cloned from a hepatoma sample. The fragment contained the putative 147 bp cellular exon in which HBV integration took place (Dejean et al., 1986).

Four positive 3' co-terminal clones were isolated from the  $2 \times 10^6$  plaques screened and the restriction maps were deduced for each of the cDNA clone EcoRI inserts. The longest one was identified lambda-13. The restriction map of lambda-13 is shown in Fig. 1.

Referring to Fig. 1, the insert of clone lambda-13 is nearly a full-length cDNA for the hap gene. Noncoding sequences (lines) and coding sequences (boxed portion) are indicated. Restriction sites are:

R EcoRI  
Bg BglII  
M MaeI  
X XhoI  
K KpnI  
P PvuII  
B BamHI  
H HindIII.

The lambda-13 clone was subjected to nucleotide sequence analysis. The nucleotide sequence is shown in Fig. 2. The nucleotide sequence of the hap cDNA is presented in the 5' to 3' orientation. The numbers on the right refer to the position of the nucleotides. Numbers above the deduced translated sequence indicate amino acid residues. The four short open reading frames in the 5' untranslated region are underlined. Adenosine residues (20) are found at the 3' end of lambda-13. The putative polyadenylation signal site (AATAAA) is boxed. The region homologous to the DNA-binding domain of known thyroid/steroid hormone receptors is indicated by horizontal arrows. The exon, previously cloned from a HCC sample genomic DNA library and in which HBV integration took place, is bracketed.

This invention of course includes variants of the nucleotide sequence shown in Fig. 2 encoding hap protein or a serotypic variant of hap protein exhibiting the same immunological reactivity as hap protein.

The DNA sequence of the invention is in a purified form. Generally, the DNA sequence is free of human serum proteins, viral proteins, and nucleotide sequences encoding these proteins. The DNA sequence of the invention can also be free of human tissue.

The DNA sequence of the invention can be used as a probe for the detection of a nucleotide sequence in a biological material, such as tissue or body fluids. The polynucleotide probe can be labeled with an atom or inorganic radical, most commonly using a radionuclide, but also perhaps with a heavy metal.

In some situations it is feasible to employ an antibody which will bind specifically to the probe hybridized to a single stranded DNA or RNA. In this instance, the antibody can be labeled to allow for detection. The same types of labels which are used for the probe can also be bound to the antibody in accordance with known techniques.

Conveniently, a radioactive label can be employed. Radioactive labels include  $^{32}\text{P}$ ,  $^3\text{H}$ ,  $^{14}\text{C}$ , or the like. Any radioactive label can be employed, which provides for an adequate signal and has sufficient half-life. Other labels include ligands, that can serve as a specific binding member to a labeled antibody, fluorescers, chemilumescers, enzymes, antibodies which can serve as a specific binding pair member for a labeled ligand, and the like. The choice of the label will be governed by the effect of the label on the rate of hybridization and binding of the probe to the DNA or RNA. It will be necessary that the label provide sufficient sensitivity to detect the amount of DNA or RNA available for hybridization.

Ligands and anti-ligands can be varied widely. Where a ligand has a natural receptor, namely ligands such as biotin, thyroxine, and cortisol, these ligands can be used in conjunction with labeled naturally occurring receptors. Alternatively, any compound can be used, either haptenic or antigenic, in combinations with an antibody.

Enzymes of interest as labels are hydrolases, particularly esterases and glycosidases, or oxidoreductases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, etc. Chemilumescers include luciferin and luminol.

#### A.2. Amino Acid Sequence of Protein Encoded by hap Gene

Based upon the sequence of the hap cDNA, the amino acid sequence of the protein encoded by hap gene was

determined. With reference to Fig. 2, the deduced amino acid sequence encoded by the gene reveals a long open reading frame of 448 amino acids corresponding to a predicted polypeptide of relative molecular mass 51,000.

A putative initiator methionine codon and an in-frame terminator codon are positioned respectively at nucleotides 322 and 1666 in the sequence (Fig. 2). Despite that two other methionine codons are found 4 and 26 triplets downstream from the first ATG, the first one is the initiation codon of translation.

The coding sequence is preceded by a 5' region of at least 321 nucleotides which contains four short open reading frames delineated by initiator and stop codons (Fig. 2). Translation usually starts, in eukaryotes, at the 5' most ATG triplet, but the finding of open reading frames in the 5' 'untranslated' region is not unprecedented (Kozak, 1986). It is not known yet whether those sequences are used for translation and exert any function in the cell.

In the 3' untranslated region, 1326 nucleotides long, no long open reading frame is present. A putative polyadenylation signal (AATAAA) is found 19 bp upstream from the polyadenylation site.

It will be understood that the present invention is intended to encompass the protein encoded by the hap gene, i. e. hap protein, and fragments thereof in highly purified form. The hap protein can be expressed in a suitable host containing the DNA sequence of the invention. This invention also includes polypeptides in which all or a portion of the binding site of hap protein is linked to a larger carrier molecule, such as a polypeptide or a protein, and in which the resulting product exhibits specific binding *in vivo* and *in vitro*. In this case, the polypeptide can be smaller or larger than the proteinaceous binding site of the protein of the invention.

It will be understood that the polypeptide of the invention encompasses molecules having equivalent peptide sequences. By this it is meant that peptide sequences need not be identical. Variations can be attributable to local mutations involving one or more amino acids not substantially affecting the binding capacity of the polypeptide. Variations can also be attributable to structural modifications that do not substantially affect binding capacity. Thus, for example, this invention is intended to cover serotypic variants of hap protein.

Three particular regions of hap gene are of interest. Two of them are located in the D region (amino acids comprised between 46 and 196), which have been shown by the inventors to be highly immunogenic. Amino acids 46-196 have the sequence:

**GlnHisArgHisThrAlaGlnSerIleGluThrGlnSerThrSerSerGluGlu  
LeuValProSerProProSerProLeuProProProArgValTyrLysProCysPheValCys  
GlnAspLysSerSerGlyTyrHisTyrGlyValSerAlaCysGluGlyCysLysGlyPhePhe  
ArgArgSerIleGlnLysAsnMetIleTyrThrCysHisArgAspLysAsnCysValIleAsn  
LysValThrArgAsnArgCysGlnTyrCysArgLeuGlnLysCysPheGluValGlyMetSer  
LysGluSerValArgAsnAspArgAsnLysLysLysLysGluThrSerLysGlnGluCysThr  
GluSerTyrGluMetThrAlaGluLeuAspAspLeuThrGluLysIleArgLysAlaHisGln  
GluThrPheProSerLeuCys.**

One peptide of interest in the D region is comprised of acids 151-167 and has the sequence:

**ValArgAsnAspAspAsnLysLysLysLysGluThrSerLysGlnGluCys.**

A second peptide in the D region is located between amino acids 175 and 185. This peptide has the amino acid sequence:

**AlaGluLeuAspAspLeuThrGluLysIleArg.**

Another peptide of interest is located at the end of C region between amino acids 440 and 448. This peptide has the amino acid sequence:

**GlyValSerGlnSerProLeuValGln.**

Other peptides having formulas derived from the nucleotide sequence of hap gene can be used as reagents,



particularly to obtain antibodies for diagnostic purposes, as defined hereinabove.

The most favorable region is found in the hinge region (amino acids 147 to 193). This region includes amino acids 150 to 170, corresponding to the following criteria:

- 5 - The region includes very hydrophilic sequences, namely, the sequences 154-160 (No. 1/Hopp); 155-161 (No. 1/Doolittle); 155-159 (No. 1/acrophilic).
- The region includes a peptide, namely, amino acids 156-162, No. 5 in mobility.
- The polypeptide of this region has a low probability of adopting a structure in the form of a folded sheet or a helix, but, in contrast, a good probability of an omega loop and one beta-turn, very marked in the Asp-Arg-Asn-Lys tetrapeptide.
- 10 - The region does not have a potential site of N-glycosylation nearby; several suggestions in this zone can be made:

**Val-Arg-Asn-Asp-Arg-Asn-Lys-Lys-Lys-Lys-Glu-Thr-Ser-Lys-**

15 **Gln-Glu-Cys (peptide 1);**

Peptide 1 corresponds to amino acids 151-167 and permits finding Cys 167, which is present in the sequence and enables attachment to a carrier (it will be noted that this peptide corresponds to a consensus sequence of phosphorylation by kinase A).

20 Peptide 1 can be shortened by N-turn while preserving the beta-turn and by C-turn while replacing Ser by Cys to maintain the possibility of coupling at this level:

25 **Asn-Asp-Arg-Asn-Lys-Lys-Lys-Lys-Glu-Thr-Cys (peptide 2).**

Peptide 2 is also favorable, but is clearly less favorable than Peptide 1 from the viewpoint of hydrophilicity as of its higher potential for spatial organization (probably as amphiphilic helix).

Finally, it will be noted that the C-terminal end constitutes a preferred region as a function of its mobility, but it nevertheless remains very hydrophobic. For example, the following peptide is contemplated:

30 **Cys-Gly-Val-Ser-Gln-Ser-Pro-Leu-Val-Gln (peptide 3).**

Peptide 3 can be fixed in a specific manner by an N-terminal Cys in such a way as to reproduce its aspect on the protein.

The nucleotide sequences of hap gene encoding those peptides are as follows:

35 For peptide 1:

**GTCAGGAATGACAGGAACAAGAAAAAGAAGGAGACTTCGAAGCAAGAATGC.**

For peptide 2:

40 **GGGGTCACTCAGTCACCACTCGTGCAA.**

For peptide 3:

**AATGACAGGAACAAGAAAAAGAAGGAGACT.**

45 For peptide of amino acids 175-185:

**GCTGAGTTGGACCATCTCACAGAGAAGATTCCGA.**

50 The polypeptides of the invention can be injected in mice, and monoclonal and polyclonal antibodies can be obtained. Classical methods can be used for the preparation of hybridomas. The antibodies can be used to quantify the amount of human receptors produced by patients in order to correlate the pathological states of illness and quantity of receptors or the absence of such receptors.

55 Epitope-bearing polypeptides, particularly those whose N-terminal and C-terminal amino acids are free, are accessible by chemical synthesis using techniques well known in the chemistry of proteins. For example, the synthesis of peptides in homogeneous solution and in solid phase is well known.

In this respect, recourse may be had to the solid phase synthesis of peptides using the method of Merrifield, J. Am. Chem. Assoc. 85, 2149-2154 (1964) or the method of synthesis in homogeneous solution described by Houben-

weyl in the work entitled "Methoden der Organischen Chemie" (Methods of Organic Chemistry), edited by E. WUNSCH, vol. 15-I and II, THIEME, Stuttgart (1974).

This method of synthesis consists of successively condensing either the successive amino acid in pairs in the appropriate order, or successive peptide fragments previously available or formed and containing already several aminoacyl residues in the appropriate order, respectively. Except for the carboxyl and amino groups which will be engaged in the formation of the peptide bonds, care must be taken to protect beforehand all other reactive groups borne by these aminoacyl groups and fragments. However, prior to the formation of the peptide bonds, the carboxyl groups are advantageously activated according to methods well known in the synthesis of peptides. Alternatively, recourse may be had to coupling reactions bringing into play conventional coupling reagents, for instance of the carbodiimide type such as 1-ethyl-3-(3-dimethyl-aminopropyl)-carbodiimide. When the amino acid group carries an additional amino group (e.g. lysine) or another acid function (e.g. glutamic acid), these groups may be protected by carbobenzoxy or t-butyloxycarbonyl groups, as regards the amino groups, or by t-butylester groups, as regards the carboxylic groups. Similar procedures are available for the protection of other reactive groups. For example, SH group (e.g. in cysteine) can be protected by an acetamidomethyl or paramethoxybenzyl group).

In the case of progressive synthesis, amino acid by amino acid, the synthesis preferably starts by the condensation of the C-terminal amino acid with the amino acid which corresponds to the neighboring aminoacyl group in the desired sequence and so on, step by step, up to the N-terminal amino acid. Another preferred technique that can be relied upon is that described by R.D. Merrifield in "Solid Phase Peptide Synthesis" (J. Am. Chem. Soc., 45, 2149-2154). In accordance with the Merrifield process, the first C-terminal amino acid of the chain is fixed to a suitable porous polymeric resin by means of its carboxylic group, the amino group of said amino acid then being protected, for example, by a t-butyloxycarbonyl group.

When the first C-terminal amino acid is thus fixed to the resin, the protective group of the amino group is removed by washing the resin with an acid, i.e. trifluoroacetic acid when the protective group of the amino group is a t-butyloxycarbonyl group.

Then the carboxylic group of the second amino acid, which is to provide the second aminoacyl group of the desired peptide sequence, is coupled to the deprotected amino group of the C-terminal amino acid fixed to the resin. Preferably, the carboxyl group of this second amino acid has been activated, for example by dicyclohexylcarbodiimide, while its amino group has been protected, for example by a t-butyloxycarbonyl group. The first part of the desired peptide chain, which comprises the first two amino acids, is thus obtained. As previously, the amino group is then deprotected, and one can further proceed with the fixing of the next aminoacyl group and so forth until the whole peptide sought is obtained.

The protective groups of the different side groups, if any, of the peptide chain so formed can then be removed. The peptide sought can then be detached from the resin, for example, by means of hydrofluoric acid, and finally recovered in pure form from the acid solution according to conventional procedures.

Depending on the use to be made of the proteins of the invention, it may be desirable to label the proteins. Examples of suitable labels are radioactive labels, enzymatic labels, fluorescent labels, chemiluminescent labels, or chromophores. The methods for labeling proteins of the invention do not differ in essence from those widely used for labeling immunoglobulin.

### A.3. Tissue Specific mRNA Distribution

In order to study expression of the hap gene, Northern blot analysis was performed using MNT as a probe and poly(A)+ RNA extracted from various human tissues and cell lines. The results are shown in Figure 3.

More particularly, Northern blot analyses were performed with poly(A)+ RNAs (15 ug per lane) extracted from different human organs and cell lines. A control hybridization with a mouse beta-actin cDNA probe is shown below the hybridizations in Fig. 3. Hap mRNA in different tissues is shown in Fig. 4A as follows:

Lane a ovary  
Lane b uterus  
Lane c HBL 100 mammary cells  
Lane d adult spleen  
Lane e 18 weeks fetal spleen  
Lane f K562  
Lane g HL60 hematopoietic cell lines  
Lane h prostatic adenoma  
Lane i kidney  
Lane j adult liver  
Lane k 18 weeks fetal liver.  
Lanes a-k correspond to a one day exposure.

Fig. 3 shows that two RNA species of 3 kb and +2.5 kb (the size of this smaller mRNA is slightly variable from one organ to another) were expressed at low abundance in ovary (lane a), uterus (lane b), HBL 100 mammary cells (lane c), adult and fetal spleen (lane d and e, respectively), and K562 and HL60 hematopoietic cell lines (lanes f and g, respectively). Surprisingly, an approximately tenfold higher level of expression was detected in prostatic adenoma (lane h) and kidney (lane i). By contrast, a single mRNA of 3000 nucleotides, expressed at low levels, was present in poly (A)<sup>+</sup> RNA from adult and fetal liver tissues (lanes j and k). Therefore, the cloned hap cDNA is likely to be a full-length copy of this transcript.

The finding of two mRNA species overexpressed in prostate and kidney, as well as the presence of a single mRNA expressed at low level in adult and fetal livers shown that hap expression is differentially regulated in those organs. This tissue specific expression provides some indication that prostate and kidney, as well as liver, could be key tissues and that hap functions in those cell types may differ.

Fig. 4 shows hap mRNA in HCC and HCC derived cell-lines as follows:

Lane a, normal liver (four days autoradiography);

Lanes b, c, d: three HCC samples (Lane b, patient Ca; Lane c, patient Mo; Lane d, patient TC1);

Lanes e, f, g: three HCC-derived cell lines (Lane e, PLC/PRF/5; Lane f, HEPG2; Lane g, HEP 3B).

The lanes b-g correspond to a one day exposure. Once again, a control hybridization with a normal beta-actin cDNA probe is shown below the hybridizations.

With reference to Fig. 4, the smaller 2.5 kb mRNA was undetectable, even after long exposure, in three adult and two fetal human livers analyzed (Fig. 4, Lane a). This differential expression in normal livers may suggest a distinct role of hap in this particular tissue.

Northern blot analysis of human HCCs and hepatoma cell lines showed almost constant alterations in hap transcription. There are two possible alternatives to explain this result. The smaller mRNA species can be simply expressed as a consequence of the cellular dedifferentiation. The tumorous liver cell, having lost its differentiated characteristics, would behave as any other cell type and thus express the same 2.5 kb mRNA as found in non-hepatic cells. However, the inability to detect such a smaller transcript in fetal livers does not seem to favor this hypothesis. On the contrary, the presence of the smaller transcript may have preceded the tumorigenesis events and would rather reflect a preneoplastic state. The presence of an inappropriately expressed hap protein, normally absent from normal hepatocytes, may have directly participated to the hepatocellular transformation. In this respect, the previous study reporting a HBV integration in the hap gene of a human HCC (Dejean et al., 1986) strongly supports the idea that hap could be causatively involved in liver oncogenesis. Indeed, in this tumor, a chimeric gene between the viral pre-S1 gene and hap may have resulted in the over-expression of a truncated hap protein. At present, it is the one found in non-hepatic tissues.

#### A.4. Expression of hap in Hepatocellular Carcinoma

Hap was first identified in a human primary liver cancer. Encouraged by this finding, poly(A)<sup>+</sup> RNA from seven hepatoma and hepatoma-derived cell lines were analyzed by Northern-blotting. Five of them contained integrated HBV DNA sequences. In addition to the 3 kb long mRNA found in normal adult and fetal liver, an additional +2.5 kb RNA species was observed, in equal or even greater amount, in three out of four HCC (Fig. 4, Lanes b, c, d) and in the PLC/PRF/5, HEPG2 and HEP3B hepatoma cell-lines (Lanes e, f, g). The size of the smaller transcript was variable from sample to sample. In addition, the two transcripts were strikingly overexpressed, at least ten fold, in the PLC/PRF/5 cells.

To test the possibility that the inappropriate expression of hap in those six tumors and tumorous cell-lines might be the consequence of a genomic DNA alteration, Southern-blotting of cellular DNA was performed using, as two probes, the MNT fragment together with a 1 kb EcoRI fragment corresponding to the 5' extremity of the cDNA insert (Fig. 2). No rearrangement and/or amplification was detected with any of these two probes which detect a different single exon (data not shown), suggesting that the hap gene was not altered at the genomic level. It is yet unknown whether the +2.5 kb mRNA, present in the liver tumorous samples and cell lines, corresponds to the same smaller transcript as that found in non-hepatic tissues. However, its presence in the liver seems to be clearly associated to the hepatocellular transformed state.

#### A.5. Hormone-binding Assay

Amino-acid homologies between the hap protein and the c-erbA/steroid receptors support the hypothesis that hap may be a receptor for a thyroid/steroid hormone-related ligand. The ability to express functional receptors *in vitro* from cloned c-erbA/steroid receptor genes led to the use of an *in vitro* translation assay to identify a putative hap ligand.

The coding region of hap was cloned into pTZ18 plasmid vector to allow *in vitro* transcription with the T7 RNA polymerase and subsequent translation in reticulocyte lysates. The results are shown in Fig. 5. More particularly, <sup>35</sup>S-

methionine-labelled products synthesized using T7 polymerase-catalysed RNA transcripts were separated on a 12% SDS-polyacrylamide gel, which was fluorographed (DMSO-PPO). The lanes in Fig. 5 are as follows:

- Lane a, pCOD 20 (sense RNA, 70 ng)  
 Lane b, PCOD 20 (140 ng)  
 Lane c, pCOD 14 (antisense RNA, 140 ng).

Figure 5 shows that the hap RNA directed the efficient synthesis of a major protein, with a 51 K relative molecular mass, consistent with the size predicted by the amino acid sequence (lanes a and b), whereas the anti-sense RNA-programmed lysate gave negligible incorporation (lane c).

Because c-erbA and hap colocalize on chromosome 3 and are more closely related according to their amino acid sequence, (<sup>125</sup>I)-T3 (triiodothyronine), -reverse T3 (3,3', 5'-triiodo-thyronine) and -T4 (thyroxine), were first tested for their binding with the in vitro translated hap polypeptide. No specific fixation with any of those three thyroid hormones could be detected. As a positive control, binding of a T3 was detected with nuclear extracts from HeLa cells. The results were negative as well when the experiment was repeated with (3H)-retinol, -retinoic acid, and -testosterone, which represent three putative ligands for hap whose receptors have not yet been cloned. Although it cannot be excluded that hap may encode a hormone independent transcriptional activator, it is more likely that hap product, i.e. the hap protein, is a receptor for a presently unidentified hormone.

#### A.6. Similarity of HAP Protein to Thyroid/Steroid Hormone Receptors

The c-erbA gene product, recently identified as a receptor for thyroid hormone (Weinberger, et al., 1986; Sap et al., 1986), as well as the steroid receptors, belong to a superfamily of regulatory proteins, which consequently to their binding with specific ligand, appear capable of activating the transcription of target genes (reviewed by Yamamoto, 1985). This activation seems to be the result of a specific binding of the hormone-receptor complex to high-affinity sites on chromatin.

Comparative sequence analysis has been made between the following different cloned steroid receptors:

- glucocorticoid receptor (GR) (Hollenberg et al., 1985; Miesfeld et al., 1986);
- oestrogen receptor (ER) (Green et al., 1986; Greene et al., 1986);
- progesterone receptor (PR) (Conneely et al., 1986; Loosfelt et al., 1986); and
- thyroid hormone receptor (c-erbA product) (Weinberger et al., 1986; Sap et al., 1986).

Mutation analysis has also been carried out. (Kumar et al., 1986; Hollenberg et al., 1987; Miesfeld et al., 1987). The results revealed the presence of two conserved regions representing the putative DNA-binding and hormone-binding domains of those molecules. It has now been discovered that hap protein is homologous to the thyroid/steroid hormone receptors.

More particularly, homology previously reported between the putative 147 bp cellular exon (bracketed in Fig. 2) and the c-erbA/steroid receptor genes led us to compare the entire hap predicted amino acid sequence with hGR, rPR, hER, and hc-erbA/thyroid hormone receptor. The five sequences have been aligned for maximal homology by the introduction of gaps. The results are depicted in Fig. 6. Specifically, the following nucleotide sequences were aligned after a computer alignment of pairs (Wilbur and Lipman, 1983):

- hap product,
- human placenta c-erbA protein (hc-erbA, Weinberger et al., 1986),
- human oestrogen receptor (hER, Green et al., 1986),
- rabbit progesterone receptor (rPR, Loosfelt et al., 1986), and
- human glucocorticoid receptor (hGR, Hollenberg et al., 1985).

A minimal number of gaps (-) was introduced in the alignment.

Amino acid residues matched in at least three of the polypeptides are boxed in Figure 6. The codes for amino acids are:

|   |     |               |
|---|-----|---------------|
| A | Ala | Alanine       |
| C | Cys | Cysteine      |
| D | Asp | Aspartic Acid |
| E | Glu | Glutamic Acid |
| F | Phe | Phenylalanine |
| G | Gly | Glycine       |
| H | His | Histidine     |

(continued)

|    |   |     |            |
|----|---|-----|------------|
|    | I | Ile | Isoleucine |
| 5  | K | Lys | Lysine     |
|    | L | Leu | Leucine    |
|    | M | Met | Methionine |
|    | N | Asn | Asparagine |
|    | P | Pro | Proline    |
| 10 | Q | Gln | Glutamine  |
|    | R | Arg | Arginine   |
|    | S | Ser | Serine     |
|    | T | Thr | Threonine  |
|    | V | Val | Valine     |
| 15 | W | Trp | Tryptophan |
|    | Y | Tyr | Tyrosine   |

The sequence comparison analysis revealed that the two regions highly conserved in the thyroid/steroid hormone receptors are similarly conserved in the hap product. Consequently, the overall organization of hap is much similar to that of the four receptors in that it can be roughly divided into four regions (arbitrarily referred to as A/B, C, D and E (Krust et al., 1986)).

In C, the most highly conserved region, extending from amino-acid 81 to 146 in hap, the nine cysteines already conserved between the four known receptors are strikingly present at the same positions. Comparison between the cysteine-rich region of hap with the corresponding region of the four receptors reveals 64% amino acid identity with hc-erbA, 59% with hER, 42% with rPR and 44% with hGR. This is schematically represented in Fig. 7.

Referring to Fig. 7, a schematic alignment of the five proteins can be seen. The division of the thyroid/steroid hormone receptor regions A/B, C, D, E is schematically represented in the hap protein. The two highly conserved regions, identified as the putative DNA-binding (region C) and hormone-binding (region E) domains of the receptors, are shown as stippled blocks. The numbers refer to the position of amino acid residues. The sequences of each of the hc-erbA product, hER, rPR and hGR receptors are compared with the hap protein. The numbers present in the stippled blocks correspond to the percentage of homology between hap protein on the one hand and each of the receptors on the other hand in the two highly conserved regions C and E. The empty blocks correspond to the non-conserved A/B and D regions.

It has also been found that hap shares 47% homology in the C region with the chicken vitamin D3 receptor (VDR), recently cloned as a partial cDNA (McDonnell et al., 1987) (data not shown). Apart from c-erbA, which contains two additional residues, the 66 amino acid long C region shows a constant length in hER, VDR, hGR, rPR and hap sequences.

Region E (residue 195-448), which is well-conserved, but to a lesser extent, shows a slightly stronger homology to hc-erbA (38%) (Fig. 7). The hap/hc-erbA homology, however, remains inferior to the identity found between hGR and rPR (90 and 51 per cent in regions C and E, respectively). No significant homology was observed when comparing the A/B (residue 1-80) and D (147-194) regions which are similarly variable, both in sequence and length, in the four known receptors.

It is thus evident from Figs. 6 and 7 that the hap product exhibits two highly homologous regions. The C domain is characterized by strikingly conserved Cys-X2-Cys units, evoking those found in the DNA-binding transcriptional factor TFIIIA (Miller et al., 1985) and in some protein that regulated development, such as Kruppel (Rosenberg et al., 1986). In the latter, the Cys-X2-Cys, together with His-X3-His units, can form metal binding fingers that are crucial for DNA-binding (Berg, 1986; Diakun et al., 1986). Similarly, the C domain of previously cloned receptors are likely to contain metal binding fingers and were shown to bind DNA (Hollenberg et al., 1987; Miesfeld et al., 1987). Since the C region of the hap gene product shares 24/66 conserved amino acids with all all steroid or thyroid hormone receptors, including all nine cysteine residues, it is likely that the hap protein is a DNA-binding protein. Hap, as c-erbA/steroid receptors, may modulate the transcription of target genes.

In addition, the the significant homology detected in the E domain suggests that hap product is a ligand-binding protein and directs the question of the nature of the putative ligand. Hap protein seems to differ too much from previously cloned hormone receptors to be a variant of one of them. In addition, the *in vitro* translated 51 K hap polypeptide failed to bind all ligands tested. Although that hap gene product could be a ligand-independent DNA-binding protein, it is believed that hap encodes a receptor for a presently unidentified circulating or intracellular ligand.

It has been proposed that steroid and thyroid hormone receptor genes were derived from a common ancestor

(Green and Chambon, 1986). This primordial gene may have provided to the receptors their common scaffolding while the hormone and target gene cellular DNA specificities were acquired through mutations accumulated in the C and E domains. Hap is both linked to the steroid receptor gene by its shorter C domain (66AA) and to the thyroid hormone receptor genes by its clearly greater homology with c-erbA in the E region (38%). This suggests that hap ligand may belong to a different hormone family.

Different functions have been assigned to the four regions defined in the glucocorticoid and oestrogen receptors (Kumar et al., 1986; Giguere et al., 1986; Miesfeld et al., 1987). By analogy, the regions C and E may represent, respectively, the putative DNA-binding and hormone-binding domains of the hap protein. The precise functions of the A/B and D domains remain unknown. The presence of the amino-terminal A/B region of the human GR has been recently shown to be necessary for full transcriptional activity (Hollenberg et al., 1987), whereas results obtained with the rat GR indicated it was dispensable (Miesfeld et al., 1987). From this alignment study it appears that hap is distinct, but closely related to the thyroid/steroid hormone receptor genes suggesting that its product may be novel ligand-dependent, DNA-binding protein.

#### A.7. Hap related genes

Southern blotting was performed on restriction enzyme-digested DNAs obtained from different organisms with labelled genomic MNT fragment containing the first exon of the cysteine-rich region of hap. The results are shown in Fig. 8. More particularly, hap related genes in vertebrates (A) and in humans (B and C) were compared. Cellular DNA (20 ug) from various sources was digested with BglII and subjected to Southern blot analysis using the MNT probe under non-stringent hybridization and washing conditions. The lanes in Fig. 8A are identified as follows:

- Lane a human liver
- Lane b domestic dog liver
- Lane c woodchuck (*marmota monax*)
- Lane d mouse liver (BALB/c strain)
- Lane e chicken erythrocytes
- Lane f cartilaginous fish (*Torpedo*).

As illustrated in Fig. 8A, BglII fragments that anneal effectively with MNT probe under non-stringent hybridization and washing conditions are present in digests of DNA from several mammals (mouse, woodchuck, dog) as well as from bird and fish. If this blotting experiment is performed at high stringency, no hybridization is observed with heterologous DNA (data not shown). These data suggest that the hybridizing sequences represent evolutionarily conserved homologs of hap.

The existence of multiple c-erbA and GR genes (Jansson et al., 1983; Weinberger et al., 1986; Hollenberg et al., 1985) encouraged a search for hap related genes in the human genome. Thus, human liver DNA digested by PstI, BamHI, and EcoRI was analyzed by Southern blot, using the MNT probe, under stringent conditions. The results are shown in Fig. 8B. After digestion of liver DNA by PstI (lane a), BamHI (lane b) or EcoRI (lane c) a single band is observed with the MNT probe in high stringency hybridization.

The same blot was hybridized with the MNT probe under non-stringent hybridization and washing conditions. The results are shown in Fig. 8C. When Southern blotting was performed under relaxed hybridization conditions, additional bands were observed in the products of each enzyme digestion (Fig. 8C, lanes a, b, c). For example, seven faint hybridizing fragments of 1, 1.7, 2.4, 3.8, 5.5, 6, 7.4 kb were observed in the BamHI digestion (lane b). None of those bands cross-hybridized with a human c-erbA probe (data not shown). A minimum of three faint bands in the PstI lane suggests the existence of at least four related hap genes in the human genome.

From a panel of somatic cell hybrids, hap was assigned to chromosome 3 (Dejean et al., 1986). To find out whether the hap related genes were all chromosomally linked or not, DNAs from human liver LA.56U and 53K cell-lines (two mouse/human somatic cell hybrids containing, altogether, most human chromosomes except chromosome 3 (Nguyen Van Cong et al., 1986)), and mouse lymphoid cells were BamHI digested, transferred to nitrocellulose, and hybridized to the MNT probe in low-stringency conditions. Of the seven faint bands present in the human liver DNA track, two at least were conserved in the LA.56U and/or L.53K cell lines DNAs digestion (data not shown) indicating that some of the hap genes do not localize on chromosome 3. Altogether the results suggest that hap belongs to a multigene family consisting of at least four members dispersed in the human genome.

The experimental procedures used in carrying out this invention will now be described in greater detail.

#### A.8. EXPERIMENTAL PROCEDURES

##### A.8.1. cDNA Cloning and Screening

Briefly, the cDNA was synthesized using oligo dT primed poly-A+ liver mRNA, using the method of Gubler and

Hoffman (1983) (C. de Taisne, unpublished data). cDNA's were size selected on a sucrose gradient and the fraction corresponding to a mean size of 3 kb was treated with EcoRI methylase. After addition of EcoRI linkers, the cDNA was digested by EcoRI and ligated to an EcoRI restricted lambda-NM1149. After in vitro encapsidation, the phages were amplified on C600 hfl and 2.10<sup>6</sup> recombinant were plated at a density of 10,000 per dish. The dishes were transferred to nylon filters and hybridized to the 350 bp EcoRI-EcoRI genomic fragment (MNT) previously described (Dejean et al., 1986). Four positive clones were isolated and the restriction map of each insert was determined. The longest one, clone lambda-13, was subjected to nucleotide sequence analysis.

#### A.8.2. Nucleotide Sequence

Clone lambda-13 DNA was sonicated, treated with the Klenow fragment of DNA polymerase plus deoxyribonucleotides (2h, 15°C) and fractionated by agarose gel electrophoresis. Fragments of 400-700 bp were excised and electroeluted. DNA was ethanol-precipitated, ligated to dephosphorylated SmaI cleaved M13 mp8 replication form DNA and transfected into Escherichia coli strain TG-1 by the high-efficiency technique of Hanahan (1983). Recombinant clones were detected by plaque hybridization using either of the four EcoRI fragments of cDNA insert as probes (Fig. 1). Single-stranded templates were prepared from plaques exhibiting positive hybridization signals and were sequenced by the dideoxy chain termination procedure (Sanger et al., 1977) using buffer gradient gels (Biggin et al., 1983).

#### A.8.3. Northern Blot

Cytoplasmic RNA was isolated from the fresh tissue using guanidine thiocyanate, and the RNA cell line was extracted using isotonic buffer and 0.5% SDS, 10 mM Na acetate pH 5.2. RNAs were then treated with hot phenol. Poly (A)+ RNA (15 µg) of the different samples were separated on a 1% agarose gel containing glyoxal, transferred to nylon filters and probed using the nick-translated MNT fragment. The experimental procedure is described in Maniatis et al. (1982).

#### A.8.4. Southern Blot

20 µg of genomic DNA was digested to completion, fractionated on a 0.8% agarose gel and transferred to nylon paper. Low stringency hybridization was performed as follows: 24 h prehybridization in 35% formamide, 5x Denhardt, 5x SSC, 300 µg/ml denatured salmon sperm DNA, at 40°C; 48 h hybridization with 35% formamide, 5x Denhardt, 5x SSC, 10% Dextran sulfate, 2.10<sup>6</sup> cpm/ml denatured <sup>32</sup>P labelled DNA probe (specific activity 5.10<sup>8</sup>cpm/µg). Washes were made in 2x SSC, 0.1 SDS, 55°C for 15 min. High stringency hybridization conditions were the same except that 50% formamide was used with 24 h hybridization. Washing was in 0.1x SSC, 0.1 SDS, 55°C for 30 min.

#### A.8.5. Construction of Plasmids for In-Vitro Translation

The 3 kb insert of phage lambda-13 was excised from the phage DNA by partial EcoRI digestion, electroeluted and digested by BamHI and HindIII. To remove most of the untranslated sequences, the 1.8 kb cDNA fragment obtained was then partially digested by MaeI (Boehringer). The 1.4 kb MaeI-MaeI fragment, extending from the first to the third MaeI site in the cDNA insert sequence (Fig. 1) and containing the complete coding region was mixed with SmaI cut dephosphorylated pTZ18 (Pharmacia), the extremities were filled in using Klenow fragment of DNA Poll (Amersham) and ligated. Two plasmids were derived: pCOD20 (sense) and pCOD14 (antisense).

#### A.8.6. Translation and hormone binding assays

pCOD20 and pCOD14 were linearized with HindIII. Capped mRNA was generated using 5 µg of DNA, 5 µM rNTP, 25 mM DTT, 100 U RNasin (Promega), 50 U T7 Pol (Genofit) in 40 mM Tris pH 8, 8mM MgCl<sub>2</sub>, 2 mM spermidine, 50 mM NaCl, in 100 µl at 37°C. Capping was performed by omitting GTP and adding CAP (m<sup>7</sup>G (5') ppp (5') G) (Pharmacia) for the 15 first minutes of the reaction. Translation was performed using rabbit reticulocyte lysate (Amersham) under the suggested conditions using 40 µl of lysate for 2.5 µg of capped RNA.

The thyroid hormone binding assays included 5 µl of lysate in (0.25 M sucrose, 0.25 KCl, 20 mM Tris (pH 7.5), 1 mM MgCl<sub>2</sub>, 2 mM EDTA, 5 mM DTT) with 1 mM <sup>125</sup>I T<sub>4</sub>, <sup>125</sup>I T<sub>3</sub> or <sup>125</sup>I rT<sub>3</sub> (specific activity: T<sub>4</sub>, rT<sub>3</sub> 1400 mCi/mg Amersham, T<sub>3</sub> 3000 mCi/mg NEN). After at least 2 h of incubation at 0°C, free was separated from bound by filtration through millipore HAWP 02500 filters using 10 ml of ice cold buffer. For testosterone, retinol, retinoic acid 10 µl of lysate was added to 45 lambda of 20 mM Tris pH 7.3, 1 mM EDTA, 50 mM NaCl, 2 mM beta-mercaptoethanol and 5 mM testosterone, 400 mM retinol or 15 mM retinoic acid (81 Ci/mmol; 60 Ci/mmol; 46 Ci/mmol; Amersham). After an overnight incubation at 0°C free was separated from bound by Dextran coated charcoal (0.5% Norit A - 0.05% T70) and

centrifugation. All experiments were performed in duplicates and parallel experiments were performed with 100 fold excess corresponding cold hormone.

## B. DIFFERENTIAL EXPRESSION AND LIGAND REGULATION OF THE RETINOIC ACID RECEPTOR $\alpha$ AND $\beta$ GENES

The recent cDNA cloning of several nuclear hormone receptors, including the steroid and thyroid hormone receptors, has revealed that their overall structures were strikingly similar. In particular, two highly conserved regions have been shown to correspond to the DNA- and hormone-binding domains (for review see Evans, 1988).

Analysis of a hepatitis B virus integration site in a human hepatocellular carcinoma led to the identification of a putative genomic exon highly homologous to the DNA-binding domain of other members of this nuclear receptor multigene family (Dejean et al., 1986). Two different cDNAs homologous to this sequence have recently been cloned (Giguere et al., 1987; Petkovich et al., 1987; de The et al., 1987) and their translation products identified as retinoic acid receptors (designated RAR  $\alpha$  and RAR  $\beta$ ) (Giguere et al., 1987; Petkovich et al., 1987; Brand et al., 1988). The two receptors have almost identical DNA- and hormone-binding domains but differ in their N-terminal part. Their respective genes map to different chromosomes, 17q21.1 for RAR  $\alpha$  (Mattei et al., 1988) and 3p24 for RAR  $\beta$  (Mattei et al., 1988), and their nucleotide sequences are only distantly related. Both genes are found in most species (Brand et al., 1988 and de The, unpublished results), suggesting an early gene duplication. Analysis of the RA-dependent gene transactivation also showed that the ED 50 of RAR  $\alpha$  and  $\beta$  were significantly different ( $10^{-8}$  and  $10^{-9}$  M, respectively), indicating that RAR- $\beta$  may mediate activation of transcription at RA concentrations 10-fold lower than those necessary for activation by RAR  $\alpha$  (Brand et al., 1988).

The existence of two different retinoic acid receptors raises a number of questions as to the biological consequences of the RAR gene publication. In particular, differences in the mechanisms of regulation or spatial expression patterns of the two receptors could account for distinct physiological roles. The tissue distribution of the transcripts for RAR  $\alpha$  and  $\beta$  and their response to RA have been studied. The results show clear differences in the spatial patterns of expression and indicate that the  $\beta$ , but not the  $\alpha$ , RAR gene is transcriptionally upregulated by RA in a protein synthesis-independent fashion. The discovery of differential expression of the RAR  $\alpha$  and  $\beta$  genes, coupled with a selective regulation of RAR  $\beta$  gene expression by RA, may prove to be important components of retinoic acid physiology. These findings strongly suggest that the two receptors are differentially involved in the various biological effects of RA. The results obtained in the study are summarized below.

The RAR  $\alpha$  gene, which is transcribed as two mRNA species of 3.2 and 2.3 kb, is overexpressed in the haematopoietic cell-lines and has an otherwise low level-expression in all the other human tissues examined. By contrast, the RAR  $\beta$  gene exhibits a much more varied expression pattern. Indeed, the two transcripts, 3 and 2.5 kb, show large variations in their levels of expression which range from undetectable (haematopoietic cell-lines) to relatively abundant (kidney, cerebral cortex, etc.). Run-on studies with the hepatoma cell-lines show that, at least in some tissues, these differences may be due to an increase in the transcription rate of the RAR  $\beta$  gene. These findings point to complex regulatory mechanisms of RAR gene expression that may confer the cells with various sensitivities to RA.

The availability of cloned RAR cDNAs prompted an investigation of possible regulation of these receptor mRNAs by RA. Exposure of hepatoma cells to RA led to a rapid increase in the level of RAR  $\beta$  transcripts, while the abundance of RAR  $\alpha$  transcripts remain unaffected. The stimulation of expression of RAR  $\beta$  mRNAs was induced by physiological concentrations of RA in a dose-dependent manner. Such autoregulation is a general feature of hormonal systems and has been shown to take place at the mRNA and protein levels, in the case of the nuclear receptors for glucocorticoids (down-regulation, Okrent et al., 1986) or vitamin D3 (up-regulation, McDonnell et al., 1987). The RA-induced upregulation of the RAR  $\beta$  transcripts was observed in the presence of protein synthesis inhibitors. *In vitro* nuclear transcript run-on assays show that the RA-induced increase in RAR  $\beta$  mRNAs levels is the consequence of an enhanced transcription. These findings demonstrate that the RAR  $\beta$  gene is transcriptionally unregulated by the RA and provide the first identification of a primary target gene for RA. The cloning of the promoter sequences of the RAR  $\beta$  gene should allow the identification of the upstream genomic elements implicated in RA responsiveness. The use of these sequences will provide a useful tool to determine which one of the  $\alpha$  and/or the  $\beta$  receptor is involved in regulating  $\beta$  RAR gene expression.

The haematopoietic cell-line HL60 has been widely used as a model for RA-induced differentiation (Strickland and Mahdavi, 1978). The data from this invention suggest that in this system RAR  $\alpha$  must be responsible for the RA-induced differentiated phenotype, since HL60 does not appear to have any RAR  $\beta$  mRNAs. Note in this respect that Davies et al. (1985) studying the RA-dependent transglutaminase expression in these cells have found an ED 50 of  $5 \times 10^{-8}$  M consistent with a RAR  $\alpha$ -mediated transactivation.

The upregulation of the  $\beta$  receptor gene by RA may have very important implications in developmental biology. Morphogen gradients are frequently implicated in cell commitment (Slack, 1987). One example of this phenomenon is the polarization of the chick limb bud where RA, the suspected morphogen, forms a concentration gradient across the



anterior-posterior axis of the developing bud (Thaller and Eichele, 1987). However, the small magnitude of this gradient (2.5 fold) is puzzling and suggests the existence of amplification mechanisms (Robertson, 1987). Since transactivation of target genes is dependent upon both receptor and ligand concentrations, a small increase in RA may result in a disproportionately larger RAR  $\beta$  effect. The effect of this RA gradient could be potentiated by a corresponding gradient in RAR  $\beta$  receptors as a consequence of upregulation by RA itself.

### B.1. Tissue distribution of the $\alpha$ and $\beta$ RAR mRNAs

To study the differential expression of the RAR  $\alpha$  and  $\beta$  genes, Northern blot analysis was performed using 5  $\mu$ g (microgram) of poly(A) + RNA extracted from various human tissues and cell-lines. A RAR  $\beta$  clone previously identified (de The et al., 1987) was used to isolate a partial cDNA clone for RAR  $\alpha$  from a hepatoma cell-line cDNA library, and the two cDNA inserts were used as probes. More particularly, poly(A) + mRNA (5  $\mu$ g) from different human tissues and cell-lines was denatured by glyoxal, separated on a 1.2% agarose gel, blotted onto nylon filters and hybridized to an  $\alpha$  (Fig. 9, upper panel) then a  $\beta$  (Fig. 9, middle panel) RAR cDNA single-stranded probe (see materials and methods, *infra*). Exposure time was 36 h. The filters were subsequently hybridized to a  $\beta$  actin probe (Fig. 9, lower panel) to ensure that equal amounts of RNA were present in the different lanes. The following abbreviations are used in Fig. 9. Sp. cord: spinal cord. C. cortex: cerebral cortex. K562 and HL60 are two haematopoietic cell-lines. PLC/PRF/5 is a hepatoma derived cell-line.

Referring to Fig. 9, the spatial distribution patterns were clearly distinct between the two receptors. The RAR  $\alpha$  probe hybridized to two transcripts of 3.2 and 2.3 kilobases (kb) with an approximately equal intensity. The two mRNAs were present at low levels in all tissues examined but were overexpressed in the haematopoietic cell-lines, K562 and HL60.

When the same filters were hybridized with the RAR  $\beta$  probe, a much more variable transcription pattern was observed (Figure 9). Two mRNA species of 3 kb and 2.5 kb were visible in most tissues, except in the spinal cord and the liver (adult or fetal) where the smaller transcript was undetectable. Major quantitative differences in the level of expression of the two transcripts were noted. The tissues examined could be classified into four groups with respect to expression of  $\beta$  receptor mRNAs: high (kidney, prostate, spinal cord, cerebral cortex, PLC/PRF/5 cells), average (liver, spleen, uterus, ovary), low (breast, testis) and undetectable (K562 and HL60 cells). The use of a  $\beta$  probe that did not hybridize to  $\alpha$ , allowed us to correct our previous description of the  $\beta$  RAR transcripts in these haematopoietic cell-lines (de The et al., 1987). The suppression of  $\beta$  receptor gene expression, associated with an overexpression of RAR  $\alpha$  mRNAs seems to be a general feature of haematopoietic cell-lines, since similar results were obtained when we repeated the study using six other cell-lines (HEL, LAMA, U937, KGI, CCRF, Burkitt) (data not shown).

### B.2. RA-induced mRNA regulation

To investigate whether retinoic acid modulates the expression of its own receptor, PLC/PRF/5 cells were grown in the presence of various concentrations of RA for different times, and RAR  $\alpha$  and  $\beta$  mRNAs were analysed by Northern blot hybridization. More particularly, semi-confluent cells were grown for 6 hr in charcoal stripped medium and retinoic acid was then added to the medium at various concentrations ( $10^{-10}$  M to  $10^{-6}$  M) for 4 hr. Control cells were treated with ethanol (E). Northern-blotting was performed as described in connection with Figure 9, except that 30  $\mu$ g of total RNA was used. Dose-response is shown in Fig. 10A.

Another analysis was performed as in Fig. 10A, except that  $10^{-6}$ M RA was used for various times (0-12 h). Time-response is shown in Fig. 10B. Exposure time was 12 hr for the  $\beta$  probe (Fig. 10B, lower panel) and four days for the  $\alpha$  probe (Fig. 10B, upper panel).

When the cells were treated with a high concentration of RA ( $10^{-6}$  M), a rapid increase in  $\beta$  receptor mRNAs was observed, and a dose-response analysis showed that this stimulatory effect was already evident at a RA concentration of  $10^{-9}$  M (Fig. 10A, lower panel). From densitometry, the magnitude of the RA-induced upregulation was 10-fold.

Since the PLC/PRF/5 cells constitutively overexpress the RAR  $\beta$  mRNAs (Fig. 9), the experiment was repeated using the HEPG2 hepatoma cell-line, which has a level of RAR  $\beta$  expression similar to that of normal adult liver (de The et al., 1987). In this case, there was a greater (50-fold) RA-induced stimulation of the levels of RAR  $\beta$  mRNAs (data not shown). Exposure of the PLC/PRF/5 cells to RA ( $10^{-6}$  M) during various periods indicated that the induction had a latency of one hour, was complete after four hours, and did not increase after an overnight treatment (Fig. 10B, lower panel). After hybridizing the same filters with an RAR  $\alpha$  probe, no variation was found in the level of the  $\alpha$  receptor mRNAs (Fig. 10, upper panel), indicating that RA had no effect on the expression of the RAR  $\alpha$  gene.

### B.3. Effect of inhibitors

To investigate the mechanism of activation of RAR  $\beta$  gene by RA, experiments with PLC/PRF/5 cells were per-

formed in the presence or absence of various inhibitors of transcription or translation, or were treated with ethanol (E) as a control.

More particularly, PLC/PRF/5 cells were exposed to charcoal stripped medium for 6 hr; subsequently ethanol (E), RA ( $10^{-6}$ M) and/or inhibitors cycloheximide (CH) 10 µg/ml or actinomycin D (AC) (5 µg/ml) were added for an additional 4 hr. Northern-blotting was carried on using 30 µg of total RNA. Figure 11 shows filters hybridized first to the RAR β probe (Fig. 11, right panel), then to the α probe (Fig. 11, left panel), and finally to a β actin probe (Fig. 11, lower panel). Exposure times were the same as for the experiments in Figure 10.

The RNA synthesis inhibitor actinomycin D (AC) abolished the RA-induced increase in the levels of RAR β transcripts (compare the RA+AC lane to the RA and E+AC lanes), while the protein synthesis inhibitor cycloheximide (CH) did not (compare lanes RA+CH to CH). Neither RA, AC, nor CH significantly affected the levels of β actin mRNA (Fig. 11, lower panel). These findings suggest that RA-induction of the β receptor gene results from a direct transcriptional effect. When the same filters were rehybridized to the RAR α probe (Fig. 11, left panel) the presence or absence of RA had no effect on the levels of RAR α mRNAs confirming that the RAR α gene is not regulated by RA.

#### B.4. Nuclear transcript elongation analysis.

Nuclear run-on experiments were carried out to determine if the enhanced expression of the RAR β gene was due to increased transcription. PRF/PLC/5 cells were grown in the presence of ethanol (E) or retinoic acid (RA), their nuclei were isolated, and transcription was performed in the presence of ( $^{32}$ P)UTP. The labelled RNAs were hybridized to filters containing single-stranded RAR β cDNA inserts in the appropriate orientation (S (sense) 10 µg and 1 µg), or in the reverse orientation (AS (antisense) 20 µg). A β actin control was also included. Exposure time was 12 hours. The results are shown in Figure 12.

The specific hybridization, which reflects the transcription rate, is clearly induced by RA. In addition, the magnitude of the increase in RAR β mRNAs is comparable when assessed by run-on assays (5 to 7 fold) or Northern analysis (8 fold to 10 fold). These experiments establish that the RAR β gene is transcriptionally upregulated by RA.

Nuclear transcript elongation assays were also used to investigate whether the higher steady-state levels of RAR β mRNAs observed in the hepatoma cells PRF/PLC/5 compared to HEPG 2 (de The et al., 1987), were related to differences in transcription rates. Transcript elongation assays were performed with PRF/PLC/5 and HEPG2 cells as described below in material and methods, in the absence of added RA. The filters contained, respectively, 10 µg and 20 µg of sense (S) and antisense (AS) RAR β cDNA inserts. Exposure time was 24 hours. The results are shown in Figure 13.

A much greater specific hybridization signal, relative to the β actin control, was observed in PRF/PLC/5 cells compared to the HEPG 2 cells (Fig. 13), indicating that their transcription rates are different. This result suggests that at least some of the variations in RAR β expression in the human tissues and cell-lines (Fig. 9) might be due, in a similar manner, to differences in the transcription rates of the RAR β gene.

#### B.5. Stability of RAR mRNAs

The level of RAR β mRNAs was slightly higher after cycloheximide treatment (compare the E lane to the CH lane in Fig. 11, right panel). In the presence of RA, CH treatment caused approximately a 50-fold increase in the level of RAR β gene expression (compare lane E to RA+CH). Such superinduction by cycloheximide has been described for several genes and associated with either transcriptional or post-transcriptional mechanisms (Greenberg et al., 1986).

To determine whether RNA stabilization was involved in the induction by CH, PLC/PRF/5 cells were first stimulated for 3 hours by RA ( $10^{-6}$ M) in the presence of CH (10 µg/ml) and extensively washed with culture medium. Transcription was then blocked by addition of actinomycin D (5 µg/ml) and the level of RAR mRNAs was monitored for the next 5 hours in the presence or absence of CH. Northern-blotting was done using 30 µg of total RNA. The results are shown in Figure 14. The filters were hybridized first to the RAR β probe (Fig. 14, right panel), then to the α probe (Fig. 14, left panel), and lastly to a β actin probe (Fig. 14, lower panel). Exposure times were as in Figure 10.

Quantification of the RAR β mRNAs levels indicated that CH indeed stabilized the β transcripts, as CH increased their half-life from approximately 50 to 80 min (Fig. 14, right panel). The combined effect of increased transcription and reduced degradation may account for the synergistic effect of RA and CH on β mRNAs levels. In the case of RAR α, cycloheximide treatment caused only a slight increase in mRNAs levels and no superinduction by RA was observed (Fig. 11, left panel). In addition, the α receptor mRNAs, which have a half life of at least 5 hours, are more stable than the RAR β transcripts (Fig. 14, left panel). A pentanucleotide, ATTTA, in A/T rich 3' non-coding regions seems to mediate mRNA degradation (Shaw and Kamen, 1986). The 3.2 kb RAR α transcript has an A/T poor 3' end (38%) and contains two such motifs (Giguere et al., 1987; Petkovich et al., 1987), whereas the 3 kb RAR β mRNA has an A/T rich 3' end (68%) and four copies of ATTTA (de The et al., 1987). These findings are consistent with the differences in

RAR  $\alpha$  and  $\beta$  mRNAs stability that have been observed.

## B.6. MATERIAL AND METHODS

### B.6.1. Biological samples and cell-lines.

Human tissue samples were obtained from early autopsies and kept at  $-80^{\circ}\text{C}$  prior to extraction. The HEPG 2 and PLC/PRF/5 hepatoma cell-lines were grown in Dulbecco's modified Eagle's medium with 10% fetal calf serum, glutamine, and antibiotics, in 5%  $\text{CO}_2$ . Semiconfluent cells were treated with RA after a 6 h wash-out in charcoal stripped medium. All-trans-retinoic acid was obtained from Sigma. Cycloheximide and actinomycin D (both from Sigma) were used at concentrations of 10 and 5  $\mu\text{g/ml}$  (micrograms/milliliter), respectively.

### B.6.2. RNA preparation.

The RNA was prepared by the hot phenol procedure (Maniatis et al., 1982). Poly(A)<sup>+</sup> mRNA was prepared by oligo (dT)-cellulose chromatography. For Northern-blot analysis, total RNA (30  $\mu\text{g}$ ) or poly(A)<sup>+</sup> mRNA (5  $\mu\text{g}$ ) was denatured by glyoxal and fractionated on a 1.2% agarose gel (Maniatis et al., 1982). The nucleic acid was transferred to nylon membranes (Amersham) by blotting and attached by UV exposure plus baking.

### B.6.3. Recombinant clones.

The  $\beta$  receptor probe was a 600 bp fragment of the cDNA previously described (de The et al., 1987) extending from the 5' end to the Xho I site, corresponding to 5' untranslated region and the A/B domain. The  $\alpha$  receptor probe was a short cDNA insert that was isolated from a PLC/PRF/5 human hepatoma cell-line cDNA library generated as described (Watson and Jackson, 1986). This library was hybridized with an RAR  $\beta$ -derived probe (nucleotides 550 to 760) corresponding to the conserved DNA-binding domain of RAR  $\beta$ . A weakly hybridizing plaque was purified, sub-cloned into M13mp18, and sequenced by the dideoxy procedure. This clone was found to be identical to RAR  $\alpha$  and extended from nucleotides 358 to 587, corresponding to the C and D domains (Giguere et al., 1987). Since this cDNA insert contains some regions homologous to the RAR  $\beta$  cDNA, cross-hybridization has been occasionally observed, particularly in cell-lines that overexpress RAR  $\beta$  mRNAs.

### B.6.4. Hybridization procedure.

The two cDNA inserts were subcloned into M13 and used to generate high specific activity (greater than  $10^9$  c.p.m./ $\mu\text{g}$ ) single-stranded probes by elongation of a sequencing primer with  $^{32}\text{P}$  labelled dTTP (3000 Ci/mmol) and unlabelled nucleotides by Klenow polymerase. The resulting double-stranded DNA was digested using a unique site in the vector, fractionated on a urea/acrylamide sequencing gel, and the labelled single-stranded insert electroeluted. These probes ( $5 \times 10^6$  cpm/ml) were hybridized to the filters in 7% (w/v) sodium dodecyl sulfate (SDS), 0.5 M  $\text{NaPO}_4$  pH 6.5, 1 mM ethylenediaminetetraacetate (EDTA), and 1 mg/ml bovine serum albumin (BSA) at  $68^{\circ}\text{C}$  overnight. The filters were washed in 1% SDS, 50 mM NaCl, 1 mM EDTA at  $68^{\circ}\text{C}$  for 10 min and autoradiographed at  $-70^{\circ}\text{C}$  using Kodak XAR films and intensifying screens. A mouse  $\beta$  actin probe was used to rehybridize the filters and check that all lanes contained equal amounts of RNA.

### B.6.5. Nuclear run-on experiments.

Nuclear transcript elongation assays were performed as described (Mezger et al., 1987). PLC/PRF/5 or HEPG 2 cells ( $10^8$ ) were challenged with ethanol or with  $10^{-6}$  M RA for 6 hours in charcoal-stripped medium. After isolation of the nuclei, transcription was performed in a final volume of 100  $\mu\text{l}$  (microliters) with 150 uCi (microcuries) of ( $\alpha^{32}\text{P}$ ) UTP (3000 Ci/mmol). Typical incorporation ranged between 2 and  $6 \times 10^7$  cpm. The labelled RNA was hybridized to nylon filters (Amersham) containing 10  $\mu\text{g}$  and 1  $\mu\text{g}$  of a 3' end RAR  $\beta$  cDNA insert (position 2495 to 2992, de The et al., 1987) cloned in M13; 20  $\mu\text{g}$  of the same insert in the reverse orientation were included as a negative control. A plasmid containing a mouse  $\beta$  actin insert (4  $\mu\text{g}$ ) provided a positive and quantitative hybridization control. Hybridization was performed with a probe concentration of  $2 \times 10^7$  cpm/ml for 48 hours.

The relative intensity of hybridization signals in Northern-blotting and run-on experiments was estimated using a Hoefer scanning densitometer and the appropriate computer program.

Our results showing a direct autoregulation of the transcription of the RAR- $\beta$  gene implies that the retinoic acid receptor  $\beta$  binds to its own gene promoter sequences. To identify those sequences, several 5' coterminal RAR- $\beta$  cDNA clones were derived from the PRF/PLC/5 library previously described. Nucleotide sequence analysis showed that these

clones extended our previous  $\lambda$  13 RAR- $\beta$  clone by 72 bp, which are shown in Figure 15. Thus, this invention also provides the 72 bp nucleotide sequence shown in Fig. 15, as well as a cloned DNA sequence encoding a polypeptide of hap gene, wherein the sequence has the formula

```

5                                     CCCATGC
GAGCTGTTTGAAGGACTGGGATGCCGAGAACGCGAGCGATCCGAGCAGGGTTTGTCTGGGCACCGT
ATGTTTGACTGTATGGATGTTCTGTCAAGTGAGTCCTGGGCAAACTCTGATTCTACACTGCGAGTCC
10 GTCTTCCTGCATGCTCCAGGAGAAAGCTCTCAAAGCATGCTTCAGTGGATTGACCCAAACCGAATC

GCAGCATCGGCACACTGCTCAATCAATTGAAACACAGAGCACCAGCTCTGAGGAACTCGTCCCAAG
CCCCCATCTCCACTTCCTCCCCCTCGAGTGATCAAACCTGCTTCGTCTGCCAGGACAAATCATC
AGGGTACCACTATGGGGTCAGCGCTGTGAGGGATGAAGGGCTTTTCCGCAGAAGTATTCAGAAG
20 AATATGATTTACACTTGTCAACGAGATAAGAAGTGTGTTATTAATAAAGTCACCAGGAATCGATGC
CAATACTGTGACTCCAGAAGTGCTTTGAAGTGGGAATGTCCAAAGAATCTGTCAGGAATGACAGG
AACAAGAAAAAGAAGGAGACTTCGAAGCAAGAATGCACAGAGAGCTATGAAATGACAGCTGAGTTG
25 GACGATCTCACAGAGAAGATCCGAAAAGCTCACCAGGAAACTTTCCCTTCACTCTCGCAGCTGGGT
AAATACACCACGAATTCAGTGCTGACCATCGAGTCCGACTGGACCTGGGCCTCTGGGACAAATTC
30 AGTGAAGTGGCCACCAAGTGCATTATTAAGATCGTGGAGTTTGCTAAACGTCTGCCTGGTTTCACT
GGCTTGACCATCGCAGACCAAATTACCCTGCTGAAGGCCGCTGCCTGGACATCCTGATTCTTAGA
ATTTGCACCAGGTATACCCAGAACAAAGACACCATGACTTTCTCAGACGGCCTTACCCTAAATCGA
35 ACTCAGATGCACAATGCTGGATTTGGTCTCTGACTGACCTTGTGTTACCTTTGCCAACCAGCTC
CTGCCTTTGGAAATGGATGACACAGAAAACAGGCCTTCTCAGTGCCATCTGCTTAATCTGTGGAGAC
CGCCAGGACCTTGAGGAACCGACAAAAGTAGATAAGCTACAAGAACCATTGCTGGAAGCACTAAAA
40 ATTTATATCAGAAAAAGACGACCCAGCAAGCCTCACATGTTTCCAAAGATCTTAATGAAAAATCACA
GATCTCCGTAGCATCAGTGCTAAAGGTGCAGAGCGTGTAATTACCTTGAAAATGGAAATTCCTGGA
45 TCAATGCCACCTCTCATTCAAGAAATGATGGAGAATTCTGAAGGACATGAACCCCTTGACCCCAAGT
TCAAGTGGGAACACAGCAGAGCACAGTCTAGCATCTCACCAGCTCAGTGGAAAACAGTGGGGTC
AGTCAGTCACCACTCGTGCAATAA,
50

```

and serotypic variants thereof, wherein said DNA is in a purified form.

This 72 bp sequence was used as a probe to screen a human genomic library. Six overlapping clones were derived, and a 6 kb HindIII - BamHI insert containing the probe was subcloned into PTZ 16 at the same sites to give rise to the plasmid pPROHAP. Since this genomic DNA insert is limited by the BamHI site present in the original  $\lambda$  13 clone and contains the additional 72 bp of the 5' end of the mRNA, it also contains the promoter region and all the elements necessary for the RAR- $\beta$  gene expression and regulation. Preliminary S1 analysis using the plasmid pPROHAP end labelled at the BamHI site suggest that the cloned RAR- $\beta$  cDNA are full-size and that the cap site is indeed located in the 90 bp BamHI-EcoRI fragment.

A complete restriction map of the HindIII-BamHI genomic DNA insert is shown in Figure 16.

Plasmid pPROHAP was transfected into the E. coli strain DH5 $\alpha$ F' (from B.R.L.). A viable culture of E. coli strain DH5 $\alpha$ F' transformed with plasmid pPROHAP was deposited on November 29, 1988, with the National Collection of Cultures of Microorganisms or Collection Nationale de Cultures de Micro-organismes (C.N.C.M.) of Institut Pasteur, Paris, France, under Culture Collection Accession No. C.N.C.M. I-821.

This DNA insert, which is characterized by its restriction map and partial nucleotide sequence (or some of its fragment), provides a tool to assess RAR- $\beta$  function, because it must contain a RAR responsive enhancer. Several constructs in which this promoter region controls the expression of indicator genes, such as the  $\beta$ -galactosidase or the chloramphenicol acetyl transferase (CAT), have been designed. Transient or stable expression, in eucaryotic cells, of these constructs, together with an expression vector of RAR- $\beta$ , provides a useful model system to directly assess stimulation of RAR- $\beta$  by a retinoid.

Thus, this invention also provides a recombinant DNA-molecule comprising a DNA sequence of coding for a retinoic acid receptor, said DNA sequence coding on expression in a unicellular host for a polypeptide displaying the retinoic acid and DNA binding properties of RAR- $\beta$  and being operatively linked to an expression control sequence in said DNA molecule.

It should be apparent that the foregoing techniques as well as other techniques known in the field of medicinal chemistry can be employed to assay for agonists and antagonists of ligand binding to RAR- $\beta$  and binding of the RAR- $\beta$  protein to DNA. Specifically, this invention makes it possible to assay for a substance that enhances the interaction of the ligand, the RAR- $\beta$  protein, and the DNA, or combinations of these materials to elicit an observable or measurable response. The substance can be an endogenous physiological substance or it can be a natural or synthetic drug.

This invention also makes it possible to assay for an antagonist that inhibits the effect of an agonist, but has no biological activity of its own in the RAR- $\beta$  effector system. Thus, for example, the invention can be employed to assay for a natural or synthetic substance that competes for the same receptor site on the RAR- $\beta$  protein or the DNA that the agonist occupies, or the invention can be employed to assay for a substance that can act on an allosteric site, which may result in allosteric inhibition.

It will be understood that this invention is not limited to assaying for substances that interact only in a particular way, but rather the invention is applicable to assaying for natural or synthetic substances, which can act on one or more of the receptor or recognition sites, including agonist binding sites, competitive antagonist binding sites (accessory sites), and non-competitive antagonist or regulatory binding sites (allosteric sites).

A convenient procedure for carrying out the method of the invention involves assaying a system for stimulation of RAR- $\beta$  by a retinoid. For example, as a retinoid binds to the receptor, the receptor-ligand complex will bind to the responsive promoter sequences and will activate transcription. For example, transcription of the  $\beta$ -galactosidase or CAT genes can be determined. The method of this invention makes it possible to screen  $\beta$ -receptor binding retinoids. In addition, this invention makes it possible to carry out blood tests for RAR- $\beta$  activity in patients.

In summary, a hepatitis B virus (HBV) integration in a 147 bp cellular DNA fragment homologous to steroid receptors and c-erbA/thyroid hormone receptor genes previously isolated from a human hepatocellular carcinoma (HCC) was used as a probe to clone the corresponding complementary DNA from a human liver cDNA library. The nucleotide sequence analysis revealed that the overall structure of the cellular gene, named hap, is similar to that of DNA-binding hormone receptors. That is, it displays two highly conserved regions identified as the putative DNA-binding and hormone-binding domains of the c-erbA/steroid receptors. Six out of seven hepatoma and hepatoma-derived cell-lines express a 2.5 kb hap mRNA species which is undetectable in normal adult and fetal livers but present in all non-hepatic tissues analyzed. Low stringency hybridization experiments revealed the existence of hap related genes in the human genome. Taken together, the data suggest that the hap product may be a member of a new family of ligand-responsive regulatory proteins whose inappropriate expression in liver seems to correlate with the hepatocellular transformed state.

Because the known receptors control the expression of target genes that are crucial for cellular growth and differentiation, an altered receptor could participate in the cell transformation. In that sense, avian v-erbA oncogene, which does not by itself induce neoplasms in animals, potentiates the erythroblast transformant effect of v-erbB and other oncogenes of the src family (Kahn et al., 1986). It has been shown that the v-erbA protein has lost its hormone-binding potential (Sap et al., 1986), presumably as a result of one or several mutations it has accumulated in its putative ligand-binding domain. It has been also suggested (Edwards et al., 1979) that the growth of human breast tumors are correlated to the presence of significant levels of ER. This invention may provide a novel example in which a DNA-binding protein would again relate to the oncogenic transformation by interfering with the transcriptional regulation of target genes. DNA-transfection assays using the native hap cDNA as well as 'altered' hap genes derived from various HCC can provide important information concerning any transforming capacity.

Following is a more detailed identification of the literature citations appearing above in parenthesis:

Beasley, R.P., and Hwang, L.Y. (1984). Epidemiology of Hepatocellular Carcinoma In *Viral Hepatitis and Liver Disease*, G.N. Vyas, J.L. Dienstag, J.H. Hoofnagle, eds, (Grune and Stratton, Inc). pp. 209-224.

- 55

- complementary DNA encoding the avian receptor for vitamin D. *Science*, 235, 1214-1217.
- Mattei, M.G., Petkovich, M., Mattei, J.F., Brand, N. and Chambon, P. *Hum. Genet.* (In the press).
- Mattei, M.G., de The, H., Mattei, J.F., Marchio, A., Tiollais, P. and Dejean, A. *Hum. Genet.* (In the press).
- McDonnell, D.P., Mangelsdorf, D.J., Pike, J.W., Haussler, M.R. and O'Malley, B.M. (1987) *Science* 235, 1214-1217.
- 5 Mezger, V., Bensaude, O., and Morange, M. (1987) *Dev. Biol.* 124, 544-550.
- Miesfeld, R., Rusconi, S., Godowski, P.J., Maler, B.A., Okret, S., Wilkström, A.C., Gustafsson, J.A. and Yamamoto, K.R. (1986). Genetic complementation of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. *Cell*, 46, 389-399.
- 10 Miesfeld, R., Godowski, P. J., Maler, B. A. and Yamamoto, K. R. (1987). Glucocorticoid receptor mutants that define a small region sufficient for enhancer activation. *Science*, 236, 423-427.
- Miller, J., McLachlan, A.D. and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *The EMBO J.*, 4, 1609-1614.
- Nguyen Van Cong, Weil, D., Finaz, C., Cohen-Haguenaue, O., Gross, M.S., Jegou-Foubert, C., de Tand, M.F., Cochet, C., de Grouchy, J., Frezal, J. (1986). Panel of twenty-five independent man-rodent hybrids for human
- 15 genetic marker mapping. *Ann. Genet.*, 29, 20-26.
- Okrent, S., Poellinger, L., Dong, Y., and Gustafsson, J.A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5899-5903.
- Petkovich, M., Brand, N.J., Krust, A. and Chambon, P. (1987) *Nature* 330, 444-450.
- Roberts, A.B. and Sporn, M.B. (1984) in *The retinoids* Vol. 2, eds. Robertson, M. (1987) *Nature* 330, 420-421.
- Rosenberg, U.G., Schröder, C., Preiss, A., Kienlin, A., Côté, S., Riede, I. and Jackle, H. (1986). Structural homology of the product of the *Drosophila* Krüppel gene with *Xenopus* transcription factor IIIA. *Nature*, 319, 336-339.
- 20 Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 74, 5463-5467.
- Sap, J., Munoz, A., Damm, K., Goldberg, Y., Ghysdael, J., Leutz, A., Beug, H. and Vennström, B. (1986). The c-erb-A protein is a high affinity receptor for thyroid hormone. *Nature*, 324, 635-640.
- 25 Shaw, G. and Kamen, R. (1986) *Cell* 46, 659-667.
- Slack, J.M.W. (1987) *Nature* 327, 553-554.
- Sporn, M.B., Roberts, A.B. and Goodman, D.S. (Academic Press, New York) pp. 209-286.
- Strickland, S. and Mahdavi, M. (1978) *Cell* 15, 393-403.
- Szmuness, W. (1973). Hepatocellular carcinoma and the hepatitis B virus evidence for a causal association. *Prog. Med. Virol.*, 24, 40-69.
- 30 Thaller, C. and Eichele, G. (1987) *Nature* 327, 625-628.
- de The, H., Marchio, A., Tiollais, P. and Dejean, A. (1987) *Nature* 330, 667-670.
- Tiollais, P., Pourcel, C. and Dejean, A. (1985). The hepatitis B virus. *Nature*, 317, 489-495.
- Watson, C.F. and Jackson, J.F. (1986) in *DNA Cloning Techniques: A Practical Approach*, ed. Glover, D. (IRL, Oxford), pp. 79-88.
- 35 Weinberger, C., Thompson, C.C., Ong, E.S., Lebo, R., Gruol, D.J. and Evans, R.M. (1986). The c-erb-A gene encodes a thyroid hormone receptor. *Nature*, 324, 641-646.
- Wilbur, W.J. and Lipman, D.J. (1983). Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA*, 80, 726-730.
- 40 Yamamoto, K.R. (1985). Steroid receptor regulated transcription of specific genes and gene networks. *Ann. Rev. Genet.*, 19, 209-252.

#### Claims

- 45 1. A polypeptide comprising an amino acid sequence which is named hap protein and which consists of the following amino acid sequence:
- 50
- 55

Met Phe Asp Cys Met Asp Val Leu Ser Val Ser Pro Gly Gln  
 Ile Leu Asp Phe Tyr Thr Ala Ser Pro Ser Ser Cys Met Leu  
 5 Gln Glu Lys Ala Leu Lys Ala Cys Phe Ser Gly Leu Thr Gln  
 Thr Glu Trp Gln His Arg His Thr Ala Gln Ser Ile Glu Thr  
 Gln Ser Thr Ser Ser Glu Glu Leu Val Pro Ser Pro Pro Ser  
 10 Pro Leu Pro Pro Pro Arg Val Tyr Lys Pro Cys Phe Val Cys  
 Gln Asp Lys Ser Ser Gly Tyr His Tyr Gly Val Ser Ala Cys  
 Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Lys Asn  
 15 Met Ile Tyr Thr Cys His Arg Asp Lys Asn Cys Val Ile Asn  
 Lys Val Thr Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys  
 Cys Phe Glu Val Gly Met Ser Lys Glu Ser Val Arg Asn Asp  
 Arg Asn Lys Lys Lys Lys Glu Thr Ser Lys Gln Glu Cys Thr  
 20 Glu Ser Tyr Glu Met Thr Ala Glu Leu Asp Asp Leu Thr Glu  
 Lys Ile Arg Lys Ala His Gln Glu Thr Phe Pro Ser Leu Cys  
 Gln Leu Gly Lys Tyr Thr Thr Asn Ser Ser Ala Asp His Arg  
 25 Val Arg Leu Asp Leu Gly Leu Trp Asp Lys Phe Ser Glu Leu  
 Ala Thr Lys Cys Ile Ile Lys Ile Val Glu Phe Ala Lys Arg  
 Leu Pro Gly Phe Thr Gly Leu Thr Ile Ala Asp Gln Ile Thr  
 30 Leu Leu Lys Ala Ala Cys Leu Asp Ile Leu Ile Leu Arg Ile  
 Cys Thr Arg Tyr Thr Pro Glu Gln Asp Thr Met Thr Phe Ser  
 Asp Gly Leu Thr Leu Asn Arg Thr Gln Met His Asn Ala Gly  
 Phe Gly Pro Leu Thr Asp Leu Val Phe Thr Phe Ala Asn Gln  
 35 Leu Leu Pro Leu Glu Met Asp Asp Thr Glu Thr Gly Leu Leu  
 Ser Ala Ile Cys Leu Ile Cys Gly Asp Arg Gln Asp Leu Glu  
 Glu Pro Thr Lys Val Asp Lys Leu Gln Glu Pro Leu Leu Glu  
 40 Ala Leu Lys Ile Tyr Ile Arg Lys Arg Arg Pro Ser Lys Pro  
 His Met Phe Pro Lys Ile Leu Met Lys Ile Thr Asp Leu Arg  
 Ser Ile Ser Ala Lys Gly Ala Glu Arg Val Ile Thr Leu Lys  
 45 Met Glu Ile Pro Gly Ser Met Pro Pro Leu Ile Gln Glu Met  
 Met Glu Asn Ser Glu Gly His Glu Pro Leu Thr Pro Ser Ser  
 50 Ser Gly Asn Thr Ala Glu His Ser Pro Ser Ile Ser Pro Ser  
 Ser Val Glu Asn Ser Gly Val Ser Gln Ser Pro Leu Val Gln

or serotypic variants of this sequence,

55 provided that this polypeptide displays the retinoic acid and DNA binding properties of RAR-beta.

2. Polypeptide as claimed in claim 1, which is free from human blood derived proteins, virus, viral protein, human tissues and human tissue components.



3. A polypeptide selected from the group consisting in the polypeptides (a) to (g) :

(a)

5 Gln His Arg His Thr Ala Gln Ser Ile Glu Thr Gln Ser Thr  
 Ser Ser Glu Glu Leu Val Pro Ser Pro Pro Ser Pro Leu Pro  
 Pro Pro Arg Val Tyr Lys Pro Cys Phe Val Cys Gln Asp Lys  
 10 Ser Ser Gly Tyr His Tyr Gly Val Ser Ala Cys Glu Gly Cys  
 Lys Gly Phe Phe Arg Arg Ser Ile Gln Lys Asn Met Ile Tyr  
 Thr Cys His Arg Asp Lys Asn Cys Val Ile Asn Lys Val Thr  
 Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys Cys Phe Glu  
 15 Val Gly Met Ser Lys Glu Ser Val Arg Asn Asp Arg Asn Lys  
 Lys Lys Lys Glu Thr Ser Lys Gln Glu Cys Thr Glu Ser Tyr  
 Glu Met Thr Ala Glu Leu Asp Asp Leu Thr Glu Lys Ile Arg  
 20 Lys Ala His Gln Glu Thr Phe Pro Ser Leu Cys

(b)

25 Val Arg Asn Asp Arg Asn Lys Lys Lys Lys Glu Thr Ser Lys  
 Gln Glu Cys (peptide 1) ;

(c)

30 Asn Asp Arg Asn Lys Lys Lys Lys Glu Thr Cys (peptide  
 2) ;

(d)

35 Cys Gly Val Ser Gln Ser Pro Leu Val Gln (peptide 3) ;

(e)

40 Ala Glu Leu Asp Asp Leu Thr Glu Lys Ile Arg

(f)

45 Met Phe Asp Cys Met Asp Val Leu Ser Val Ser Pro Gly Gln  
 Ile Leu Asp Phe Tyr Thr Ala Ser Pro Ser Ser Cys Met Leu  
 Gln Glu Lys Ala Leu Lys Ala Cys Phe Ser Gly Leu Thr Gln  
 50 Thr Glu Trp Gln His Arg His Thr Ala Gln Ser

(g)

55 His Glu Pro Leu Thr Pro Ser Ser Ser Gly Asn Thr Ala Glu  
 His Ser Pro Ser Ile Ser Pro Ser Ser Val Glu Asn Ser Gly  
 Val Ser Gln Ser Pro Leu Val Gln

or any serotypic variant of one of the said polypeptides (a) to (g), the polypeptide or the serotypic variant having

binding capacity with retinoic acid.

4. A cloned DNA sequence comprising a sequence encoding for a polypeptide as claimed in claims 1 and 2, the said cloned sequence comprising the following formula (a1) :

5

10

15

20

25

30

35

40

45

50

55

ATG TTT GAC TGT ATG GAT GTT CTG TCA GTG AGT CCT GGG CAA  
 ATC CTG GAT TTC TAC ACT GCG AGT CCG TCT TCC TGC ATG CTC  
 CAG GAG AAA GCT CTC AAA GCA TGC TTC AGT GGA TTG ACC CAA  
 ACC GAA TGG CAG CAT CGG CAC ACT GCT CAA TCA ATT GAA ACA  
 CAG AGC ACC AGC TCT GAG GAA CTC GTC CCA AGC CCC CCA TCT  
 CCA CTT CCT CCC CCT CGA GTG TAC AAA CCC TGC TTC GTC TGC  
 CAG GAC AAA TCA TCA GGG TAC CAC TAT GGG GTC AGC GCC TGT  
 GAG GGA TGT AAG GGC TTT TTC CGC AGA AGT ATT CAG AAG AAT  
 ATG ATT TAC ACT TGT CAC CGA GAT AAG AAC TGT GTT ATT AAT  
 AAA GTC ACC AGG AAT CGA TGC CAA TAC TGT CGA CTC CAG AAG  
 TGC TTT GAA GTG GGA ATG TCC AAA GAA TCT GTC AGG AAT GAC  
 AGG AAC AAG AAA AAG AAG GAG ACT TCG AAG CAA GAA TGC ACA  
 GAG AGC TAT GAA ATG ACA GCT GAG TTG GAC GAT CTC ACA GAG  
 AAG ATC CGA AAA GCT CAC CAG GAA ACT TTC CCT TCA CTC TGC  
 CAG CTG GGT AAA TAC ACC ACG AAT TCC AGT GCT GAC CAT CGA  
 GTC CGA CTG GAC CTG GGC CTC TGG GAC AAA TTC AGT GAA CTG  
 GCC ACC AAG TGC ATT ATT AAG ATC GTG GAG TTT GCT AAA CGT

CTG CCT GGT TTC ACT GGC TTG ACC ATC GCA GAC CAA ATT ACC  
 CTG CTG AAG GCC GCC TGC CTG GAC ATC CTG ATT CTT AGA ATT  
 TGC ACC AGG TAT ACC CCA GAA CAA GAC ACC ATG ACT TTC TCA  
 GAC GGC CTT ACC CTA AAT CGA ACT CAG ATG CAC AAT GCT GGA  
 TTT GGT CCT CTG ACT GAC CTT GTG TTC ACC TTT GCC AAC CAG  
 CTC CTG CCT TTG GAA ATG GAT GAC ACA GAA ACA GGC CTT CTC  
 AGT GCC ATC TGC TTA ATC TGT GGA GAC CGC CAG GAC CTT GAG  
 GAA CCG ACA AAA GTA GAT AAG CTA CAA GAA CCA TTG CTG GAA  
 GCA CTA AAA ATT TAT ATC AGA AAA AGA CGA CCC AGC AAG CCT  
 CAC ATG TTT CCA AAG ATC TTA ATG AAA ATC ACA GAT CTC CGT  
 AGC ATC AGT GCT AAA GGT GCA GAG CGT GTA ATT ACC TTG AAA  
 ATG GAA ATT CCT GGA TCA ATG CCA CCT CTC ATT CAA GAA ATG  
 ATG GAG AAT TCT GAA GGA CAT GAA CCC TTG ACC CCA AGT TCA  
 AGT GGG AAC ACA GCA GAG CAC AGT CCT AGC ATC TCA CCC AGC  
 TCA GTG GAA AAC AGT GGG GTC AGT CAG TCA CCA CTC GTG CAA  
 TAA,

or degeneracy variants of this formula encoding the said polypeptide.

- 5 5. DNA sequence as claimed in claim 4, which is free of human serum protein components, such as human tissue or serum proteins.
6. DNA sequence having any of the following formulae (b1) to (g1) corresponding respectively to the aminoacid sequences (b) to (g) of the claim 3:

(b1)

10 GTC AGG AAT GAC AGG AAC AAG AAA AAG AAG GAG ACT TCG AAG  
CAA GAA TGC;

(c1)

15 AAT GAC AGG AAC AAG AAA AAG AAG GAG ACT;

(d1)

20 GGG GTC ACT CAG TCA CCA CTC GTG CAA;

(e1)

25 GCT GAG TTG GAC CAT CTC ACA GAG AAG ATC CGA;

(f1)

30 ATG TTT GAC TGT ATG GAT GTT CTG TCA GTG AGT CCT GGG CAA  
ATC CTC GAT TTC TAC ACT GCG AGT CCG TCT TCC TGC ATG CTC  
CAG GAG AAA GCT CTC AAA GCA TGC TTC AGT GGA TTG ACC CAA  
35 ACC GAA TGG CAG CAT CGG CAC ACT GCT CAA TCA;

(g1)

40 CAT GAA CCC TTG ACC CCA AGT TCA AGT GGG AAC ACA GCA GAG  
CAC ACT CCT AGC ATC TCA CCC AGC TCA GTG GAA AAC AGT GGG  
GTC ACT CAG TCA CCA CTC GTG CAA;

or degeneracy variants of these formulae.

- 45 7. A DNA probe consisting essentially of a radio-nucleotide bonded to the DNA sequence of any of claims 4 or 5.
8. A hybrid duplex molecule consisting essentially of the DNA sequence of claims 4 or 5, hydrogen bonded to a nucleotide sequence of complementary base sequence.
- 50 9. Hybrid duplex molecule as claimed in claim 8, wherein said nucleotide sequence is either a DNA sequence, or a RNA sequence.
10. A process for selecting, from a group of nucleotide sequences, a nucleotide sequence, e.g. a DNA sequence or a RNA sequence, said nucleotide sequence being preferably labelled, e.g. by a radionucleotide, said nucleotide sequence coding for thyroid or steroid hormone receptor, or for RAR- $\beta$  or a portion thereof, said process comprising the step of determining which of said nucleotide sequences hybridizes to a DNA sequence as claimed in claims 4 to 6.

11. Process as claimed in claim 10, wherein said nucleotide sequence is selected by Southern blot technique, under high stringency conditions performed as follows: 24 h prehybridization in 50% formamide, 5x Denhardt, 5x SSC, 300 up/ml denatured salmon sperm DNA, at 40°C; 48 h hybridization with 35% formamide, 5x Denhardt, 5x SSC, 10% Dextran sulfate,  $2 \cdot 10^6$  cpm/ml denatured  $^{32}\text{P}$  labelled DNA probe (specific activity  $5 \cdot 10^8$  cpm/ug); washes in 0.1x SSC, 0.1 SDS, 55°C for 30 minutes.
12. Process as claimed in claim 10, wherein said nucleotide sequence is selected by Northern blot technique according to Maniatis, T., Fritsch, E., and Sambrook, J. (1982), Molecular cloning: a laboratory manual (Cold Spring Harbor, New-York: Cold Spring Harbor Laboratory).
13. An E. coli bacterial culture in a purified form, wherein the culture comprises E. coli cells containing DNA, wherein a portion of said DNA comprises the DNA sequence as claimed in claim 4.
14. A method for assaying a fluid for the presence of an agonist or antagonist to thyroid or steroid hormone receptor, wherein the method comprises:
  - (A) providing an aqueous solution containing a known concentration of the proteinaceous receptor as claimed in claim 1;
  - (B) incubating the receptor with the fluid suspected of containing the agonist or antagonist under conditions sufficient to bind the receptor to the agonist or antagonist; and
  - (C) determining whether there is change in concentration of the proteinaceous receptor in the aqueous solution.
15. A method for assaying a fluid for the presence of an agonist or antagonist to retinoic acid receptor RAR- $\beta$ , wherein the method comprises:
  - (A) providing an aqueous solution containing a known concentration of the proteinaceous receptor as claimed in claim 1;
  - (B) incubating the receptor with the fluid suspected of containing the agonist or antagonist under conditions sufficient to bind the receptor to the agonist or antagonist; and
  - (C) determining whether there is change in concentration of the proteinaceous receptor in the aqueous solution.
16. Method as claimed in claim 15, wherein the receptor and the agonist or antagonist form a complex.
17. Method as claimed in claim 15, wherein a cross-linking agent is present in an amount sufficient to inhibit dissociation of the receptor and the agonist or antagonist.
18. A recombinant DNA molecule comprising a DNA sequence of coding for a retinoic acid receptor, said DNA sequence coding on expression in a unicellular host for a polypeptide displaying the retinoic acid and DNA binding properties of RAR- $\beta$  and being operatively linked to an expression control sequence in said DNA molecule.
19. The recombinant DNA molecule of claim 18 wherein the DNA sequence is any one of claims 4 to 6.
20. Plasmid pPROHAP, deposited at the CNCM on November 29, 1988, under n° I-821.
21. Bacterial culture as claimed in claim 13, wherein said cells are comprised of E. coli strain DH5 $\alpha$ F', deposited at the CNCM on November 29, 1988, under n° I-821.
22. Use of polypeptides according to claims 1 to 3, as a reagent to obtain antibodies for diagnostic purposes.

#### Patentansprüche

1. Polypeptid, das eine Aminosäuresequenz umfaßt, das hap-Protein genannt wird und das aus der folgenden Aminosäuresequenz:

Met Phe Asp Cys Met Asp Val Leu Ser Val Ser Pro Gly Gln  
 Ile Leu Asp Phe Tyr Thr Ala Ser Pro Ser Ser Cys Met Leu  
 5 Gln Glu Lys Ala Leu Lys Ala Cys Phe Ser Gly Leu Thr Gln  
 Thr Glu Trp Gln His Arg His Thr Ala Gln Ser Ile Glu Thr  
 Gln Ser Thr Ser Ser Glu Glu Leu Val Pro Ser Pro Pro Ser  
 10 Pro Leu Pro Pro Pro Arg Val Tyr Lys Pro Cys Phe Val Cys  
 Gln Asp Lys Ser Ser Gly Tyr His Tyr Gly Val Ser Ala Cys  
 Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Lys Asn  
 Met Ile Tyr Thr Cys His Arg Asp Lys Asn Cys Val Ile Asn  
 15 Lys Val Thr Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys  
 Cys Phe Glu Val Gly Met Ser Lys Glu Ser Val Arg Asn Asp  
 Arg Asn Lys Lys Lys Lys Glu Thr Ser Lys Gln Glu Cys Thr  
 20 Glu Ser Tyr Glu Met Thr Ala Glu Leu Asp Asp Leu Thr Glu  
 Lys Ile Arg Lys Ala His Gln Glu Thr Phe Pro Ser Leu Cys  
 Gln Leu Gly Lys Tyr Thr Thr Asn Ser Ser Ala Asp His Arg  
 25 Val Arg Leu Asp Leu Gly Leu Trp Asp Lys Phe Ser Glu Leu  
 Ala Thr Lys Cys Ile Ile Lys Ile Val Glu Phe Ala Lys Arg  
 Leu Pro Gly Phe Thr Gly Leu Thr Ile Ala Asp Gln Ile Thr  
 30 Leu Leu Lys Ala Ala Cys Leu Asp Ile Leu Ile Leu Arg Ile  
 Cys Thr Arg Tyr Thr Pro Glu Gln Asp Thr Met Thr Phe Ser  
 Asp Gly Leu Thr Leu Asn Arg Thr Gln Met His Asn Ala Gly  
 Phe Gly Pro Leu Thr Asp Leu Val Phe Thr Phe Ala Asn Gln  
 35 Leu Leu Pro Leu Glu Met Asp Asp Thr Glu Thr Gly Leu Leu  
 Ser Ala Ile Cys Leu Ile Cys Gly Asp Arg Gln Asp Leu Glu  
 Glu Pro Thr Lys Val Asp Lys Leu Gln Glu Pro Leu Leu Glu  
 40 Ala Leu Lys Ile Tyr Ile Arg Lys Arg Arg Pro Ser Lys Pro  
 His Met Phe Pro Lys Ile Leu Met Lys Ile Thr Asp Leu Arg  
 Ser Ile Ser Ala Lys Gly Ala Glu Arg Val Ile Thr Leu Lys  
 45 Met Glu Ile Pro Gly Ser Met Pro Pro Leu Ile Gln Glu Met  
 Met Glu Asn Ser Glu Gly His Glu Pro Leu Thr Pro Ser Ser  
 Ser Gly Asn Thr Ala Glu His Ser Pro Ser Ile Ser Pro Ser  
 50 Ser Val Glu Asn Ser Gly Val Ser Gln Ser Pro Leu Val Gln

oder serotypischen Varianten dieser Sequenz besteht, mit der Maßgabe, daß dieses Polypeptid die Retinoesäure- und DNA-Bindungseigenschaften von RAR-beta zeigt.

- 55
2. Polypeptid nach Anspruch 1, das frei von aus menschlichem Blut abgeleiteten bzw. abgeleitetem Proteinen, Virus, viralem Protein, menschlichem Gewebe und menschlichen Gewebestandteilen ist.
  3. Polypeptid, das aus der Gruppe ausgewählt ist, die aus den Polypeptiden (a) bis (g):

(a)

5 Gln His Arg His Thr Ala Gln Ser Ile Glu Thr Gln Ser Thr  
 Ser Ser Glu Glu Leu Val Pro Ser Pro Pro Ser Pro Leu Pro  
 Pro Pro Arg Val Tyr Lys Pro Cys Phe Val Cys Gln Asp Lys  
 Ser Ser Gly Tyr His Tyr Gly Val Ser Ala Cys Glu Gly Cys  
 10 Lys Gly Phe Phe Arg Arg Ser Ile Gln Lys Asn Met Ile Tyr  
 Thr Cys His Arg Asp Lys Asn Cys Val Ile Asn Lys Val Thr  
 Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys Cys Phe Glu  
 Val Gly Met Ser Lys Glu Ser Val Arg Asn Asp Arg Asn Lys  
 15 Lys Lys Lys Glu Thr Ser Lys Gln Glu Cys Thr Glu Ser Tyr  
 Glu Met Thr Ala Glu Leu Asp Asp Leu Thr Glu Lys Ile Arg  
 Lys Ala His Gln Glu Thr Phe Pro Ser Leu Cys

(b)

20 Val Arg Asn Asp Arg Asn Lys Lys Lys Lys Glu Thr Ser Lys  
 Gln Glu Cys ( Peptid 1 );

(c)

25 Asn Asp Arg Asn Lys Lys Lys Lys Glu Thr Cys (Peptid  
 30 2);

(d)

35 Cys Gly Val Ser Gln Ser Pro Leu Val Gln ( Peptid 3 );

(e)

40 Ala Glu Leu Asp Asp Leu Thr Glu Lys Ile Arg

(f)

45 Met Phe Asp Cys Met Asp Val Leu Ser Val Ser Pro Gly Gln  
 Ile Leu Asp Phe Tyr Thr Ala Ser Pro Ser Ser Cys Met Leu  
 Gln Glu Lys Ala Leu Lys Ala Cys Phe Ser Gly Leu Thr Gln  
 Thr Glu Trp Gln His Arg His Thr Ala Gln Ser

(g)

50 His Glu Pro Leu Thr Pro Ser Ser Ser Gly Asn Thr Ala Glu  
 His Ser Pro Ser Ile Ser Pro Ser Ser Val Glu Asn Ser Gly  
 Val Ser Gln Ser Pro Leu Val Gln

55 oder einer jeden serotypischen Variante eines dieser Polypeptide (a) bis (g) besteht, wobei das Polypeptid oder die serotypische Variante eine Bindungsfähigkeit zu Retinoesäure besitzen.

4. Geklonte DNA-Sequenz, die eine Sequenz umfaßt, die für ein Polypeptid nach den Ansprüchen 1 und 2 codiert,

wobei die geklonte Sequenz die folgende Formel (a1):

5       ATG TTT GAC TGT ATG GAT GTT CTG TCA GTG AGT CCT GGG CAA  
       ATC CTG GAT TTC TAC ACT GCG AGT CCG TCT TCC TGC ATG CTC  
       CAG GAG AAA GCT CTC AAA GCA TGC TTC AGT GGA TTG ACC CAA  
       ACC GAA TGG CAG CAT CGG CAC ACT GCT CAA TCA ATT GAA ACA  
 10       CAG AGC ACC AGC TCT GAG GAA CTC GTC CCA AGC CCC CCA TCT  
       CCA CTT CCT CCC CCT CGA GTG TAC AAA CCC TGC TTC GTC TGC  
       CAG GAC AAA TCA TCA GGG TAC CAC TAT GGG GTC AGC GCC TGT  
       GAG GGA TGT AAG GGC TTT TTC CGC AGA AGT ATT CAG AAG AAT  
 15       ATG ATT TAC ACT TGT CAC CGA GAT AAG AAC TGT GTT ATT AAT  
       AAA GTC ACC AGG AAT CGA TGC CAA TAC TGT CGA CTC CAG AAG  
       TGC TTT GAA GTG GGA ATG TCC AAA GAA TCT GTC AGG AAT GAC  
 20       AGG AAC AAG AAA AAG AAG GAG ACT TCG AAG CAA GAA TGC ACA  
       GAG AGC TAT GAA ATG ACA GCT GAG TTG GAC GAT CTC ACA GAG  
       AAG ATC CGA AAA GCT CAC CAG GAA ACT TTC CCT TCA CTC TGC  
 25       CAG CTG GGT AAA TAC ACC ACG AAT TCC AGT GCT GAC CAT CGA  
       GTC CGA CTG GAC CTG GGC CTC TGG GAC AAA TTC AGT GAA CTG  
       GCC ACC AAG TGC ATT ATT AAG ATC GTG GAG TTT GCT AAA CGT  
 30  
       CTG CCT GGT TTC ACT GGC TTG ACC ATC GCA GAC CAA ATT ACC  
       CTG CTG AAG GCC GCC TGC CTG GAC ATC CTG ATT CTT AGA ATT  
 35       TGC ACC AGG TAT ACC CCA GAA CAA GAC ACC ATG ACT TTC TCA  
       GAC GGC CTT ACC CTA AAT CGA ACT CAG ATG CAC AAT GCT GGA  
       TTT GGT CCT CTG ACT GAC CTT GTG TTC ACC TTT GCC AAC CAG  
 40       CTC CTG CCT TTG GAA ATG GAT GAC ACA GAA ACA GGC CTT CTC  
       AGT GCC ATC TGC TTA ATC TGT GGA GAC CGC CAG GAC CTT GAG  
       GAA CCG ACA AAA GTA GAT AAG CTA CAA GAA CCA TTG CTG GAA  
       GCA CTA AAA ATT TAT ATC AGA AAA AGA CGA CCC AGC AAG CCT  
 45       CAC ATG TTT CCA AAG ATC TTA ATG AAA ATC ACA GAT CTC CGT  
       AGC ATC AGT GCT AAA GGT GCA GAG CGT GTA ATT ACC TTG AAA  
       ATG GAA ATT CCT GGA TCA ATG CCA CCT CTC ATT CAA GAA ATG  
 50       ATG GAG AAT TCT GAA GGA CAT GAA CCC TTG ACC CCA AGT TCA  
       AGT GGG AAC ACA GCA GAG CAC AGT CCT AGC ATC TCA CCC AGC  
       TCA GTG GAA AAC AGT GGG GTC AGT CAG TCA CCA CTC GTG CAA  
 55       TAA,

oder Degenerationsvarianten dieser Formel, die das Polypeptid codieren, umfaßt.

5. DNA-Sequenz nach Anspruch 4, die frei von menschlichen Serumproteinbestandteilen, wie menschlichem Gewebe oder Serumproteine, ist.
6. DNA-Sequenz mit einer jeden der folgenden Formeln (b1) bis (g1), welche jeweils den Aminosäuresequenzen (b) bis (g) von Anspruch 3:

(b1)

GTC AGG AAT GAC AGG AAC AAG AAA AAG AAG GAG ACT TCG AAG  
CAA GAA TGC;

(c1)

AAT GAC AGG AAC AAG AAA AAG AAG GAG ACT;

(d1)

GGG GTC ACT CAG TCA CCA CTC GTG CAA;

(e1)

GCT GAG TTG GAC CAT CTC ACA GAG AAG ATC CGA;

(f1)

ATG TTT GAC TGT ATG GAT GTT CTG TCA GTG AGT CCT GGG CAA  
ATC CTC GAT TTC TAC ACT GCG AGT CCG TCT TCC TGC ATG CTC  
CAG GAG AAA GCT CTC AAA GCA TGC TTC AGT GGA TTG ACC CAA  
ACC GAA TGG CAG CAT CGG CAC ACT GCT CAA TCA;

(g1)

CAT GAA CCC TTG ACC CCA AGT TCA AGT GGG AAC ACA GCA GAG  
CAC ACT CCT AGC ATC TCA CCC AGC TCA GTG GAA AAC AGT GGG  
GTC ACT CAG TCA CCA CTC GTG CAA;

oder Degenerationsvarianten dieser Formeln entsprechen.

7. DNA-Sonde, die im wesentlichen aus einem an die DNA-Sequenz nach einem der Ansprüche 4 oder 5 gebundenen Radionukleotid besteht.
8. Hybrid-Duplexmolekül, das im wesentlichen aus der DNA-Sequenz der Ansprüche 4 oder 5 besteht, welche über Wasserstoff an eine Nukleotidsequenz mit komplementärer Basensequenz gebunden ist.
9. Hybrid-Duplexmolekül nach Anspruch 8, worin die Nukleotidsequenz entweder eine DNA-Sequenz oder eine RNA-Sequenz ist.
10. Verfahren zum Selektieren einer Nukleotidsequenz, z.B. einer DNA-Sequenz oder einer RNA-Sequenz aus einer Gruppe von Nukleotidsequenzen, wobei die Nukleotidsequenz vorzugsweise markiert ist, beispielsweise durch ein Radionukleotid, diese Nukleotidsequenz für Schilddrüsen- oder Steroidhormonrezeptor oder für RAR-β oder einen Abschnitt davon codiert, wobei das Verfahren den Schritt der Bestimmung, welche Nukleotidsequenzen mit einer in den Ansprüchen 4 bis 6 beanspruchten DNA-Sequenz hybridisiert, umfaßt.
11. Verfahren nach Anspruch 10, worin die Nucleotidsequenz durch Southern-Blot-Technik unter hochgenauen Bedingungen, die wie folgt durchgeführt werden, selektiert wird: 24 h Vorhybridisierung in 50 % Formamid, 5 x Denhardt, 5 x SSC, 300 µp/ml denaturierte Salmesperma-DNA, bei 40°C; 48 h Hybridisieren mit 35 % Formamid, 5 x



Denhardt, 5 x SSC, 10 % Dextransulfat, 2 x 10<sup>6</sup> cpm/ml denaturierte <sup>32</sup>P-markierte DNA-Sonde (spezifische Aktivität 5 x 10<sup>8</sup>cpm/μg); Waschen in 0,1 x SSC, 0,1 SDS, bei 55°C, 30 Minuten.

12. Verfahren nach Anspruch 10, worin die Nucleotidsequenz durch Northern-Blot-Technik nach Maniatis, T., Fritsch, E. und Sambrook, J. (1982), Molecular cloning: a laboratory manual (Cold Spring Harbor, New-York: Cold Spring Harbor Laboratory) selektiert wird.
13. E. coli Bakterienkultur in gereinigter Form, worin die Kultur DNA enthaltende E. coli Zellen umfaßt, wobei ein Abschnitt dieser DNA die in Anspruch 4 beanspruchte DNA-Sequenz umfaßt.
14. Verfahren zur Untersuchung einer Flüssigkeit auf die Anwesenheit eines Agonisten oder Antagonisten für Schilddrüsen- oder Steroidhormonrezeptor, wobei das Verfahren umfaßt:
  - (A) das Vorsehen einer wäßrigen Lösung, welche eine bekannte Konzentration des in Anspruch 1 beanspruchten proteinartigen Rezeptors enthält;
  - (B) das Inkubieren des Rezeptors mit der Flüssigkeit, von der angenommen wird, daß sie den Agonisten oder Antagonisten enthält, unter Bedingungen, welche ausreichen, um den Rezeptor an den Agonisten oder Antagonisten zu binden, und
  - (C) das Feststellen, ob eine Konzentrationsänderung des proteinartigen Rezeptors in der wäßrigen Lösung eintritt.
15. Verfahren zur Untersuchung einer Flüssigkeit auf die Anwesenheit eines Agonisten oder Antagonisten für Retinoesäurerezeptor RAR-β, wobei das Verfahren umfaßt:
  - (A) das Vorsehen einer wäßrigen Lösung, welche eine bekannte Konzentration des in Anspruch 1 beanspruchten proteinartigen Rezeptors enthält,
  - (B) das Inkubieren des Rezeptors mit der Flüssigkeit, von der angenommen wird, daß sie den Agonisten oder Antagonisten enthält, unter Bedingungen, welche ausreichen, um den Rezeptor an den Agonisten oder Antagonisten zu binden, und
  - (C) das Feststellen, ob eine Konzentrationsänderung des proteinartigen Rezeptors in der wäßrigen Lösung eintritt.
16. Verfahren nach Anspruch 15, wobei der Rezeptor und der Agonist oder Antagonist einen Komplex bilden.
17. Verfahren nach Anspruch 15, wobei ein Vernetzungsmittel in einer Menge vorliegt, welche ausreichend ist, um eine Dissoziation des Rezeptors und des Agonisten oder Antagonisten zu verhindern.
18. Rekombinantes DNA-Molekül, das eine DNA-Sequenz umfaßt, die für Retinoesäurerezeptor codiert, wobei die DNA-Sequenz auf Expression in einem einzelligen Wirt für ein Polypeptid codiert, das die Retinoesäure- und DNA-Bindungseigenschaften von RAR-β zeigt und wirksam mit einer Expressionskontrollsequenz in dem DNA-Molekül verbunden ist.
19. Rekombinantes DNA-Molekül nach Anspruch 18, wobei die DNA-Sequenz eine jede der Ansprüche 4 bis 6 ist.
20. Plasmid pPROHAP, am 29. November 1988, bei CNCM unter der Nummer I-821 hinterlegt.
21. Bakterienkultur nach Anspruch 13, wobei die Zellen aus dem E. coli-Stamm DH5αF' zusammengesetzt sind, der bei CNCM am 29. November 1988 unter der Nummer I-821 hinterlegt wurde.
22. Verwendung der Polypeptide nach den Ansprüchen 1 bis 3 als Reagenz zum Erhalten von Antikörpern für Diagnosezwecke.

# Revendications

1. Polypeptide comprenant une séquence d'acides aminés, dénommé protéine hap et constitué de la séquence d'acides aminés suivante:

Met Phe Asp Cys Met Asp Val Leu Ser Val Ser Pro Gly Gln  
 Ile Leu Asp Phe Tyr Thr Ala Ser Pro Ser Ser Cys Met Leu  
 5 Gln Glu Lys Ala Leu Lys Ala Cys Phe Ser Gly Leu Thr Gln  
 Thr Glu Trp Gln His Arg His Thr Ala Gln Ser Ile Glu Thr  
 Gln Ser Thr Ser Ser Glu Glu Leu Val Pro Ser Pro Pro Ser  
 Pro Leu Pro Pro Pro Arg Val Tyr Lys Pro Cys Phe Val Cys  
 Gln Asp Lys Ser Ser Gly Tyr His Tyr Gly Val Ser Ala Cys  
 10 Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Lys Asn  
 Met Ile Tyr Thr Cys His Arg Asp Lys Asn Cys Val Ile Asn  
 Lys Val Thr Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys  
 Cys Phe Glu Val Gly Met Ser Lys Glu Ser Val Arg Asn Asp  
 Arg Asn Lys Lys Lys Lys Glu Thr Ser Lys Gln Glu Cys Thr  
 15 Glu Ser Tyr Glu Met Thr Ala Glu Leu Asp Asp Leu Thr Glu  
 Lys Ile Arg Lys Ala His Gln Glu Thr Phe Pro Ser Leu Cys  
 Gln Leu Gly Lys Tyr Thr Thr Asn Ser Ser Ala Asp His Arg  
 Val Arg Leu Asp Leu Gly Leu Trp Asp Lys Phe Ser Glu Leu  
 Ala Thr Lys Cys Ile Ile Lys Ile Val Glu Phe Ala Lys Arg  
 20 Leu Pro Gly Phe Thr Gly Leu Thr Ile Ala Asp Gln Ile Thr  
 Leu Leu Lys Ala Ala Cys Leu Asp Ile Leu Ile Leu Arg Ile  
 Cys Thr Arg Tyr Thr Pro Glu Gln Asp Thr Met Thr Phe Ser  
 Asp Gly Leu Thr Leu Asn Arg Thr Gln Met His Asn Ala Gly  
 Phe Gly Pro Leu Thr Asp Leu Val Phe Thr Phe Ala Asn Gln  
 25 Leu Leu Pro Leu Glu Met Asp Asp Thr Glu Thr Gly Leu Leu  
 Ser Ala Ile Cys Leu Ile Cys Gly Asp Arg Gln Asp Leu Glu  
 Glu Pro Thr Lys Val Asp Lys Leu Gln Glu Pro Leu Leu Glu  
 Ala Leu Lys Ile Tyr Ile Arg Lys Arg Arg Pro Ser Lys Pro  
 His Met Phe Pro Lys Ile Leu Met Lys Ile Thr Asp Leu Arg  
 30 Ser Ile Ser Ala Lys Gly Ala Glu Arg Val Ile Thr Leu Lys  
 Met Glu Ile Pro Gly Ser Met Pro Pro Leu Ile Gln Glu Met  
 Met Glu Asn Ser Glu Gly His Glu Pro Leu Thr Pro Ser Ser

35

Ser Gly Asn Thr Ala Glu His Ser Pro Ser Ile Ser Pro Ser  
 Ser Val Glu Asn Ser Gly Val Ser Gln Ser Pro Leu Val Gln

40

ou des variants sérotypiques de cette séquence,  
 à condition que ce polypeptide présente les propriétés de liaison de l'acide rétinolique et de l'ADN de RAR- $\beta$ .

45

2. Polypeptide selon la revendication 1, caractérisé en ce qu'il est dépourvu de protéines, de virus, de protéine virale, de tissus humains et de composants de tissus humains issus du sang humain.
3. Polypeptide choisi dans le groupe constitué des polypeptides (a) à (g) :

50

55

(a)

5 Gln His Arg His Thr Ala Gln Ser Il Glu Thr Gln S r Thr  
 Ser Ser Glu Glu Leu Val Pro Ser Pro Pro Ser Pro Leu Pro  
 Pro Pro Arg Val Tyr Lys Pro Cys Phe Val Cys Gln Asp Lys  
 Ser Ser Gly Tyr His Tyr Gly Val Ser Ala Cys Glu Gly Cys  
 10 Lys Gly Phe Phe Arg Arg Ser Ile Gln Lys Asn Met Ile Tyr  
 Thr Cys His Arg Asp Lys Asn Cys Val Ile Asn Lys Val Thr  
 Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys Cys Phe Glu  
 Val Gly Met Ser Lys Glu Ser Val Arg Asn Asp Arg Asn Lys  
 15 Lys Lys Lys Glu Thr Ser Lys Gln Glu Cys Thr Glu Ser Tyr  
 Glu Met Thr Ala Glu Leu Asp Asp Leu Thr Glu Lys Ile Arg  
 Lys Ala His Gln Glu Thr Phe Pro Ser Leu Cys

(b)

20 Val Arg Asn Asp Arg Asn Lys Lys Lys Lys Glu Thr Ser Lys  
 Gln Glu Cys (peptide 1);

(c)

25 Asn Asp Arg Asn Lys Lys Lys Lys Glu Thr Cys (peptide  
 2);

(d)

30 Cys Gly Val Ser Gln Ser Pro Leu Val Gln (peptide 3);

(e)

35 Ala Glu Leu Asp Asp Leu Thr Glu Lys Ile Arg

(f)

40 Met Phe Asp Cys Met Asp Val Leu Ser Val Ser Pro Gly Gln  
 Ile Leu Asp Phe Tyr Thr Ala Ser Pro Ser Ser Cys Met Leu  
 Gln Glu Lys Ala Leu Lys Ala Cys Phe Ser Gly Leu Thr Gln  
 Thr Glu Trp Gln His Arg His Thr Ala Gln Ser

(g)

45 His Glu Pro Leu Thr Pro Ser Ser Ser Gly Asn Thr Ala Glu  
 His Ser Pro Ser Ile Ser Pro Ser Ser Val Glu Asn Ser Gly  
 50 Val Ser Gln Ser Pro Leu Val Gln

ou tout variant séréotypique de l'un desdits polypeptides (a) à (g), le polypeptide et le variant séréotypique ayant une capacité de liaison vis-à-vis de l'acide rétinolique.

- 55 4. Séquence d'ADN clonée, comprenant une séquence codant pour un polypeptide selon les revendications 1 et 2, ladite séquence clonée comprenant la formule suivante (a1):

5 ATG TTT GAC TGT ATG GAT GTT CTG TCA GTG AGT CCT GGG CAA  
 ATC CTG GAT TTC TAC ACT GCG AGT CCG TCT TCC TGC ATG CTC  
 CAG GAG AAA GCT CTC AAA GCA TGC TTC AGT GGA TTG ACC CAA  
 ACC GAA TGG CAG CAT CGG CAC ACT GCT CAA TCA ATT GAA ACA  
 CAG AGC ACC AGC TCT GAG GAA CTC GTC CCA AGC CCC CCA TCT  
 10 CCA CTT CCT CCC CCT CGA GTG TAC AAA CCC TGC TTC GTC TGC  
 CAG GAC AAA TCA TCA GGG TAC CAC TAT GGG GTC AGC GCC TGT  
 GAG GGA TGT AAG GGC TTT TTC CGC AGA AGT ATT CAG AAG AAT  
 ATG ATT TAC ACT TGT CAC CGA GAT AAG AAC TGT GTT ATT AAT  
 15 AAA GTC ACC AGG AAT CGA TGC CAA TAC TGT CGA CTC CAG AAG  
 TGC TTT GAA GTG GGA ATG TCC AAA GAA TCT GTC AGG AAT GAC  
 AGG AAC AAG AAA AAG AAG GAG ACT TCG AAG CAA GAA TGC ACA  
 GAG AGC TAT GAA ATG ACA GCT GAG TTG GAC GAT CTC ACA GAG  
 20 AAG ATC CGA AAA GCT CAC CAG GAA ACT TTC CCT TCA CTC TGC  
 CAG CTG GGT AAA TAC ACC ACG AAT TCC AGT GCT GAC CAT CGA  
 GTC CGA CTG GAC CTG GGC CTC TGG GAC AAA TTC AGT GAA CTG  
 25 GCC ACC AAG TGC ATT ATT AAG ATC GTG GAG TTT GCT AAA CGT  
  
 30 CTG CCT GGT TTC ACT GGC TTG ACC ATC GCA GAC CAA ATT ACC  
 CTG CTG AAG GCC GCC TGC CTG GAC ATC CTG ATT CTT AGA ATT  
 TGC ACC AGG TAT ACC CCA GAA CAA GAC ACC ATG ACT TTC TCA  
 35 GAC GGC CTT ACC CTA AAT CGA ACT CAG ATG CAC AAT GCT GGA  
 TTT GGT CCT CTG ACT GAC CTT GTG TTC ACC TTT GCC AAC CAG  
 CTC CTG CCT TTG GAA ATG GAT GAC ACA GAA ACA GGC CTT CTC  
 40 AGT GCC ATC TGC TTA ATC TGT GGA GAC CGC CAG GAC CTT GAG  
 GAA CCG ACA AAA GTA GAT AAG CTA CAA GAA CCA TTG CTG GAA  
 GCA CTA AAA ATT TAT ATC AGA AAA AGA CGA CCC AGC AAG CCT  
 45 CAC ATG TTT CCA AAG ATC TTA ATG AAA ATC ACA GAT CTC CGT  
 AGC ATC AGT GCT AAA GGT GCA GAG CGT GTA ATT ACC TTG AAA  
 ATG GAA ATT CCT GGA TCA ATG CCA CCT CTC ATT CAA GAA ATG  
 50 ATG GAG AAT TCT GAA GGA CAT GAA CCC TTG ACC CCA AGT TCA  
 AGT GGG AAC ACA GCA GAG CAC AGT CCT AGC ATC TCA CCC AGC  
 TCA GTG GAA AAC AGT GGG GTC AGT CAG TCA CCA CTC GTG CAA  
 55 TAA,

ou des variants de dégénérescence de cette formule codant pour ledit polypeptide.

5. Séquence d'ADN selon la revendication 4, caractérisée en ce qu'elle est dépourvue de composants protéiques

du sérum humain, tels que des protéines du tissu ou du sérum humain.

6. Séquence d'ADN ayant l'une quelconque des formules (b1) à (g1) suivantes, correspondant respectivement aux séquences d'acides aminés (b) à (g) selon la revendication 3:

(b1)

GTC AGG AAT GAC AGG AAC AAG AAA AAG AAG GAG ACT TCG AAG  
CAA GAA TGC;

(c1)

AAT GAC AGG AAC AAG AAA AAG AAG GAG ACT;

(d1)

GGG GTC ACT CAG TCA CCA CTC GTG CAA;

(e1)

GCT GAG TTG GAC CAT CTC ACA GAG AAG ATC CGA;

(f1)

ATG TTT GAC TGT ATG GAT GTT CTG TCA GTG AGT CCT GGG CAA  
ATC CTC GAT TTC TAC ACT GCG AGT CCG TCT TCC TGC ATG CTC  
CAG GAG AAA GCT CTC AAA GCA TGC TTC AGT GGA TTG ACC CAA  
ACC GAA TGG CAG CAT CGG CAC ACT GCT CAA TCA;

(g1)

CAT GAA CCC TTG ACC CCA AGT TCA AGT GGG AAC ACA GCA GAG  
CAC ACT CCT AGC ATC TCA CCC AGC TCA GTG GAA AAC AGT GGG  
GTC ACT CAG TCA CCA CTC GTG CAA;

ou des variants de dégénérescence de ces formules.

7. Sonde d'ADN constituée essentiellement d'un radionucléotide lié à la séquence d'ADN de l'une quelconque des revendications 4 ou 5.

8. Molécule duplex hybride constituée essentiellement d'une séquence d'ADN selon la revendication 4 ou 5, liée par liaison hydrogène à une séquence nucléotidique de la séquence de bases complémentaire.

9. Molécule duplex hybride selon la revendication 8, caractérisée en ce que ladite séquence nucléotidique est soit une séquence d'ADN soit une séquence d'ARN.

10. Procédé de sélection, dans un groupe de séquences nucléotidiques, d'une séquence nucléotidique, par ex. une séquence d'ADN ou une séquence d'ARN, ladite séquence nucléotidique étant de préférence marquée, par ex. par un radionucléotide, ladite séquence nucléotidique codant pour un récepteur de l'hormone thyroïdienne ou stéroïde, ou pour le RAR- $\beta$  ou une partie de celui-ci, ledit procédé comprenant les étapes consistant à déterminer laquelle desdites séquences nucléotidiques s'hybride à une séquence d'ADN selon les revendications 4 à 6.

11. Procédé selon la revendication 10, caractérisé en ce que ladite séquence nucléotidique est choisie par la technique de Southern-blot, dans des conditions de forte stringence réalisées de la manière suivante: 24h de préhybridation

dans du formamide à 50%, 5x Denhardt, 5x SSC, 300 µg/ml d'ADN de sperme de saumon dénaturé, à 40°C; hybridation de 48 h avec du formamide à 35%, 5x Denhardt, 5x SSC, sulfate de Dextrane à 10%, 2.10<sup>6</sup> cpm/ml de sonde d'ADN dénaturée marquée au <sup>32</sup>P (activité spécifique 5.10<sup>8</sup> cpm/µg); lavage dans 0,1x SSC, 0,1% SDS, 55°C pendant 30 minutes.

5

12. Procédé selon la revendication 10, caractérisé en ce que ladite séquence nucléotidique est choisie par la technique de Northern-blot selon Maniatis, T., Fritsch, E., et Sambrook, J. (1982), Molecular cloning: a laboratory manual (Cold Spring Harbor, New-York: Cold Spring Harbor Laboratory).

10

13. Culture bactérienne de E. Coli sous forme purifiée, caractérisée en ce que la culture comprend des cellules de E. Coli contenant de l'ADN, et en ce qu'une partie dudit ADN comprend la séquence d'ADN selon la revendication 4.

14. Méthode de dosage d'un liquide pour la présence d'un agoniste ou d'un antagoniste du récepteur de l'hormone thyroïdienne ou stéroïde, caractérisée en ce que la méthode comprend:

15

(A) l'obtention d'une solution aqueuse contenant une quantité connue du récepteur protéique selon la revendication 1;

(B) l'incubation du récepteur avec le liquide soupçonné de contenir l'agoniste ou l'antagoniste, dans des conditions suffisantes pour lier le récepteur à l'agoniste ou à l'antagoniste; et

20

(C) la détermination d'un changement éventuel dans la concentration du récepteur protéique dans la solution aqueuse.

15. Méthode de dosage d'un liquide pour la présence d'un agoniste ou d'un antagoniste du récepteur de l'acide rétinique RAR-β, caractérisée en ce que la méthode comprend:

25

(A) l'obtention d'une solution aqueuse contenant une quantité connue du récepteur protéique selon la revendication 1;

(B) l'incubation du récepteur avec le liquide soupçonné de contenir l'agoniste ou l'antagoniste, dans des conditions suffisantes pour lier le récepteur à l'agoniste ou à l'antagoniste; et

30

(C) la détermination d'un changement éventuel dans la concentration du récepteur protéique dans la solution aqueuse.

16. Méthode selon la revendication 15, caractérisée en ce que le récepteur et l'agoniste ou l'antagoniste forment un complexe.

35

17. Méthode selon la revendication 15, caractérisée en ce qu'un agent de réticulation est présent dans une quantité suffisante pour inhiber la dissociation du récepteur et de l'agoniste ou de l'antagoniste.

18. Molécule d'ADN recombinant comprenant une séquence d'ADN codant pour un récepteur de l'acide rétinique, ladite séquence d'ADN codant, par expression dans un hôte unicellulaire, pour un polypeptide présentant les propriétés de liaison de l'acide rétinique et de l'ADN de RAR-β et liée de manière fonctionnelle à une séquence de contrôle d'expression dans ladite molécule d'ADN.

40

19. Molécule d'ADN recombinant selon la revendication 18, caractérisée en ce que ladite séquence d'ADN est l'une quelconque des revendications 4 à 6.

45

20. Plasmide pPROHAP, déposé auprès de la CNCM le 29 novembre 1988 sous le n° I-821.

21. Culture bactérienne selon la revendication 13, caractérisée en ce que lesdites cellules sont constituées de la souche de E. Coli DH5αF', déposée auprès de la CNCM le 29 novembre 1988 sous le n° I-821.

50

22. Utilisation des polypeptides selon les revendications 1 à 3, comme réactif pour obtenir des anticorps à des fins de diagnostic.

55

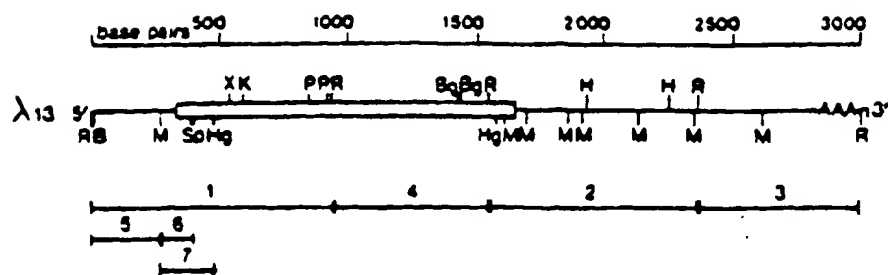


FIGURE 1

Fig. 2(a)

CCGGTLGGATCGGGAAGCATTGGAAAGGCTTTTTCGAAGCATTTACTTGGAAAGCAGAACTTCGGATCTTTCTG 75  
 CGAAGCCCCCGCGGCTGGATTGCGCGGAGCAAGCCTGCAAAATGCTAAATGATCATTTGGATCAATTACAGCC 150  
 TTTAGCTGGCTTGCTGTGATATAATTGATTCGGGCTGGCAAAAGCACCACGCTACGTGCCAAAAAAGG 225  
 CCCAGAGTTTCAATGGAGTTGGCTGGAGCTTTTCTATGCCATTTCCTCCACAGCTAGAGGATAAGCACTTTTTCAG 300  
 MetPheAspCysMetAspValLeuSerValSerProGlyGlnIleLeuAspPhe  
 ACATTGAGTGCAGGAGAGATCATGTTTGAGTGTATGGATGTTCTGTCTGCTGAGTCTGCGCAAAATCTCGCATTTIC 375  
 TyrThrAlaSerProSerSerCysMetLeuGlnGluLysAlaLeuLysAlaCysPheSerGlyLeuThrGlnThr 40  
 TACACTGGCAGTCCGCTCTTCCTGCAATCTCCAGGAGAAAGCTCTCAAGCATGCTTCAGTTCGATTGACCCAAACC 450  
 GluTrpGlnHisArgHisThrAlaGlnSerIleGluThrGlnSerThrSerSerGluGlnLeuValProSerPro 60  
 GAATGCCAGCATCGGCACACTGCTCAATCAATTGAAACACAGAGCAGCAGCTCTGAGGAAGTCTGCGAAGCCCCC 70  
 ProSerProLeuProProArgValTyrLysProCysPheValCysGlnAspLysSerSerGlyTyrHisTyr 90  
 CCAATCTGCAGCTTCCTCCCTCGAGTGTAGAAACCTGCTGCTGCTGCGCAGGACAAATCATCAGCGTACCACTAT 600



Fig 2 (b)

GlyValSerAlaCysGluGlyCysLysGlyPhePheArgSerIleGlnLysAsnMetIleTyrThrCysHis  
 100  
 CCGCTCAGCCCTGTCAGCGATGTAAAGCGCTTTTCCGAGAGTATTCAGAGAATAATGATTACAGTTGTCAC  
 110  
 120  
 ArgAspLysAsnCysValIleAsnLysValThrArgAsnArgCysGlnTyrCysArgLeuGlnLysCysPheGlu  
 130  
 CGAGATAAGAAAGTGTGTTATTAAATAAGTCCAGCAATCGATGCCAATACTGTCCACTCCAGAAAGTGCTTTGAA  
 140  
 ValGlyMetSerLysGluSerValArgAsnAspArgAsnLysLysLysGluThrSerLysGlnGluCysThr  
 150  
 GTCCGAATGTCGAAGCAATCTGTCAGGAATCAGCAGCAACAAGAAAAGAGGAGACTTCGAAGCAAGCAATGCCACA  
 160  
 170  
 GluSerTyrGluMetThrAlaGluLeuAspLeuThrGluLysIleArgLysAlaHisGlnGluThrPhePro  
 180  
 CAGAGCTATGAAATGACAGCTGAGTTGAGCGATCTCAGAGAGAGAGATCCGAAAAGCTCAGCAGGAAAACCTTTCCCT  
 190  
 200  
 SerLeuCysGlnLeuGlyLysTyrThrAsnSerSerAlaAspHisArgValArgLeuAspLeuGlyLeuTrp  
 210  
 TCAGCTCTGCCAGCTCGGTAAATACAGCAGCAATTCAGTCTGACCATCGAGTCCGACTCGAGCTCGCCCTCTCG  
 220  
 230  
 AspLysPheSerGluLeuAlaThrLysCysIleIleLysIleValGluPheAlaLysArgLeuProGlyPheThr  
 240  
 CACAAATTCACTGAAGTCCGACCAAGTCCATTATTAAAGATCGTCCAGTTTCTAAAGCTCTGCTGCTTTCACCT  
 250  
 260  
 GlyLeuThrIleAlaAspGlnIleThrLeuLeuLysAlaAlaCysLeuAspIleLeuIleLeuArgIleCysThr  
 270  
 CGCTTCACCATCCGAGACCAATTACCTGCTGAAGCCGCTGCTGCTGACATCTCATTTCTTACAAATTTCACCC  
 280  
 290  
 ArgTyrThrProGluGlnAspThrMetThrPheSerAspGlyLeuThrLeuAsnArgThrGlnMetHisAsnAla  
 300  
 AGGTATACCCCAACAAGACACCATCACTTTCTCAGACCGCCCTTACCTTAAATCGAAGTTCAGATGCCAAATGCT  
 310

Fig 2(c)

300 GlyPheGlyProLeuThrAspLeuValPheThrPheAlaAsnGlnLeuLeuProLeuGluMetAspThrGlu 1275  
 CGATTTCGTCCTCTCACTGACCTTGTCTTTCACCTTTCACCTTTCACCTTTCGAAATCGATGACACAGAA  
 320 330 340  
 ThrGlyLeuLeuSerAlaIleCysLeuIleCysGlyAspArgGlnAspLeuGluProThrLysValAspLys  
 ACAGCCCTTCTCAGTCCCATCTGCTTAAATCTCTGGAGACCGCCAGCACCTTCAGGAAACCGACAAAAGTAGATAAG 1350  
 350 360  
 LeuGlnGluProLeuLeuGluAlaLeuLysIleTyrIleArgLysArgArgProSerLysProHisMetPhePro  
 CTACAAGAACCATTCGTCGACGACGACTAAATAATTATATCAGAAAAAGACGACCCAGCAAGCCCTCAGCATGTTTCCA 1425  
 370 380 390  
 LysIleLeuMetLysIleThrAspLeuArgSerIleSerAlaLysGlyAlaGluArgValIleThrLeuLysMet  
 AAGAICTTAAATGAAAAATCAGACATCTCCGTCAGCATCAGTCGTAAGGTCCAGAGCGGTGTAATTACCTTCAAAAATG 1500  
 400 410  
 GluIleProGlySerMetProProLeuIleGlnGluMetMetGluAsnSerGluGlyHisGluProLeuThrPro  
 GAAATTCTCGCATCAATGCCACCTCTCATTCAGAAATGATGAGAAATTCAGAGCATCAACCCCTTCACCCCA 1575  
 420 430 440  
 SerSerSerGlyAsnThrAlaGluHisSerProSerIleSerProSerSerValGluAsnSerGlyValSerGln  
 AGTTCAAGTGGGAACACAGCAGACGACAGTCCTAGCATCTCACCACGCTCAGTGGGAAACAGTGGGTCAGTCAG 1650

Fig. 2 (d)

SerProLeuValGlnSTOP  
 TCACCAGCTGCTGCAATAAGACATTTTCTAGCTACTTCAACATTCGCCAGTACCTTTCAGTTCCAGGATTTAAAAAT 1725  
 GCAAGAAAAAACAATTTTCTAGCTGCTTAGTTTTCGAGCTGCAAAAGATATTTAAAGTCAAGAACGACCAAGAAAGT 1800  
 TTTCAATATGTAATCAATATAATAGCTGCTCAGCTGTGTAAGTTAGCTAGAAATACAAAGTTTTCCAAATTTTAAAAAA 1875  
 TCAGCCATTTTCATCCAACCCAGCAAACTAGTTAAAGCTTCTATTTTCTCTCTTTGAACACTCAAGATGCCATGCCAAA 1950  
 GAGCCAGTCAAAATGATTTACCCCTGCTTAAAGTTTCTGAAGACTTTGTACATACAGAAAGTATGCCCTCTCTCTT 2025  
 CTATAGCTGTAAGTTTGGTGGCTTTGCTTTTGTCTGCAIAGTCAAAAIAACCATGACACCAAGCTTATGAAATAGA 2100  
 CTACTGTACACGCTACCTAGGTTCAAAAGATAAAGTGTGTTGCTTTTCATGGAATAGTCAAGACATCAAGGTTAAG 2175  
 GAAAGACGACTATTCACAGGACTATTTGTACAGTATGCAAGATAGGCTGCAAGATATTTCTACTTTAGTTAGTTATG 2250  
 GAAGCTTGTCTTTGCTCTTTTCTGATGCTCTCAAACTGCAATCTTTTATTTTCATGTTGCCCAGTAAAGTATACAAA 2325  
 TTCCCTGCACCTAGCAGAAAGAGAAATTCCTGATCAGTGTAACTGCCAGTTTCAGTTAAATGCAATGTGATTTGTTCAT 2400  
 TGTAAATGTCACTTTAAATTAAGTGGTTTATTAAGTGTTTAATCACAATAACTACACAGTTAGTTAAAAAAAAT 2475  
 TTTTTLACAGTAAATAGCTCCAGCCAGAAACAGCTTTTTCAGTGTAAAGTTTGTGTTTGTTCACAGCC 2550  
 ATTACGCCAAATTTTCATGCCATAATTACGAGGCTGCTGACAGTCCAGCAATGTAAGTCTAGTGTCTCTCTGATT 2625  
 CATGCCGTGATATTGGGATTTTTCAGCCCTTCTTTCATGCCAAGGCTAATTATATTACATCCCAAGCAAAACAG 2700  
 GCATAGAAATCGCCCTCTTTCAGCTTGTTCATCAGTATGCAAGCAGAGTCAAGCTGTGCTAGAGTCTTAAACAG 2775  
 ATACAACTGTCAAGTTTCTTCTAGTCTCATTTAAGCAGTACTGGAATTTTGTGATATATTAGCAAGTCTGTG 2850  
 ATGTACTTTCAGTGGCTGCTGTTTGTAGCATTCAGATTTGTTTACAAATGCTTTGTATGTTTCATATAGTGTTLA 2925  
 CCTTTTTCATGGAGTCTGCTGCGCAAGATTAATAATATATTTTAAAAAAAATAAAAAA 2992

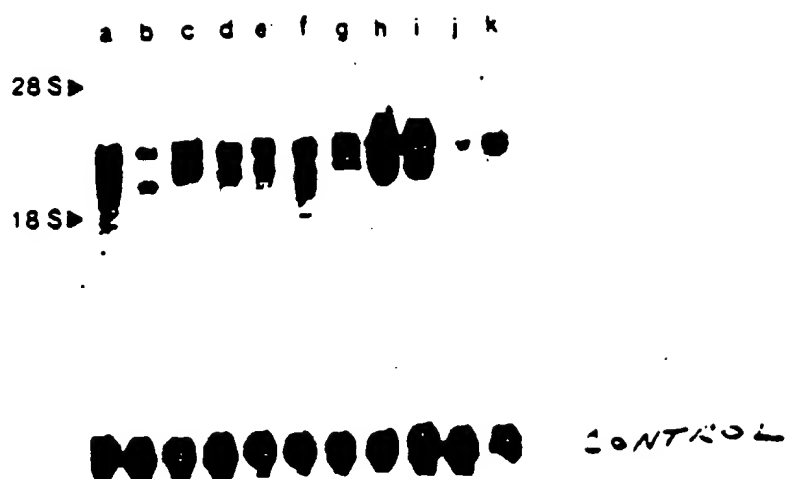


FIGURE 3

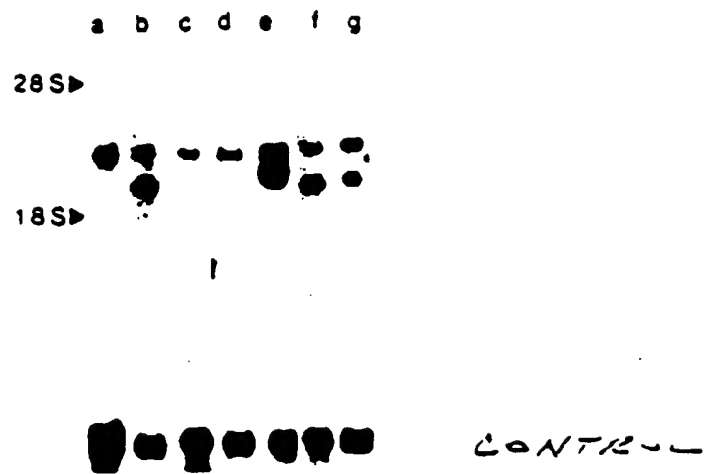


FIGURE 4

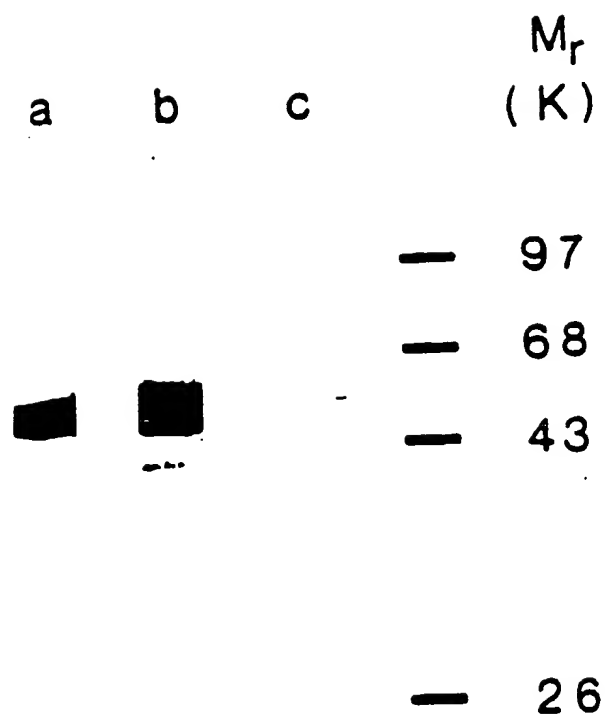


FIGURE 5

FIGURE 6

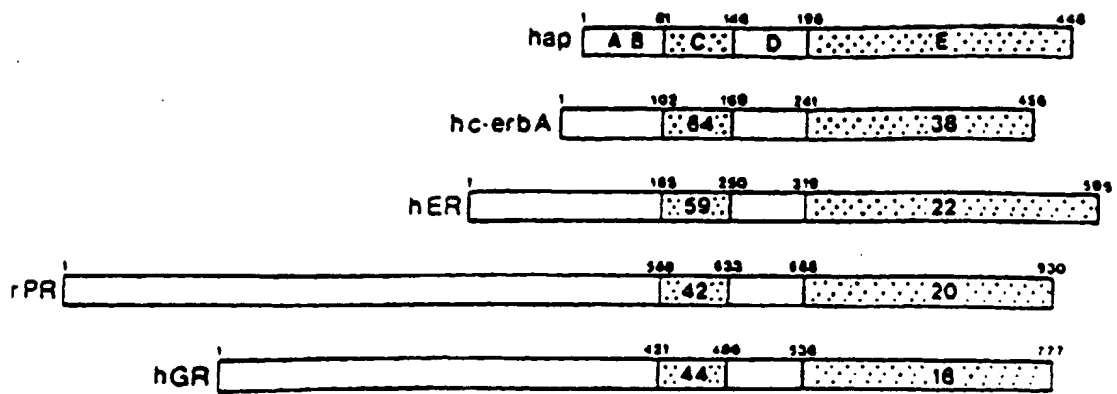


FIGURE 7



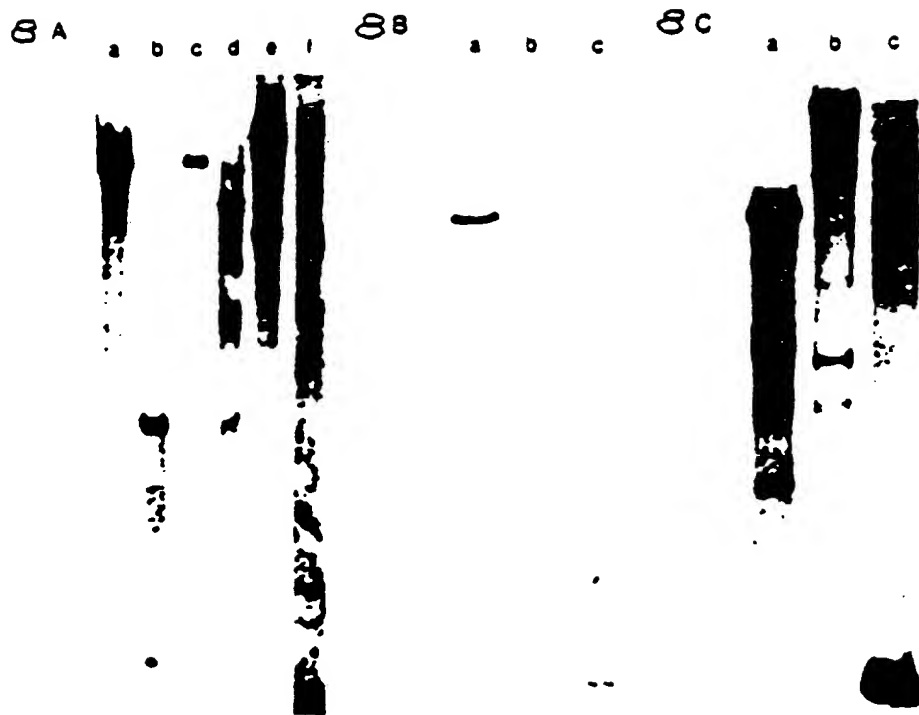


FIGURE 8

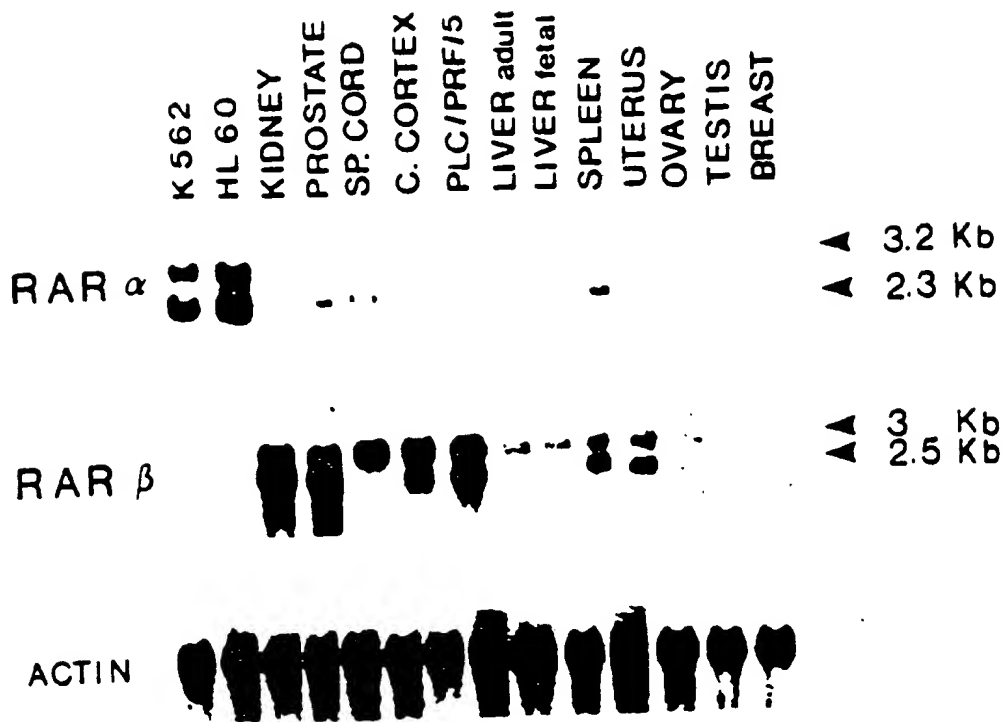


FIGURE 9

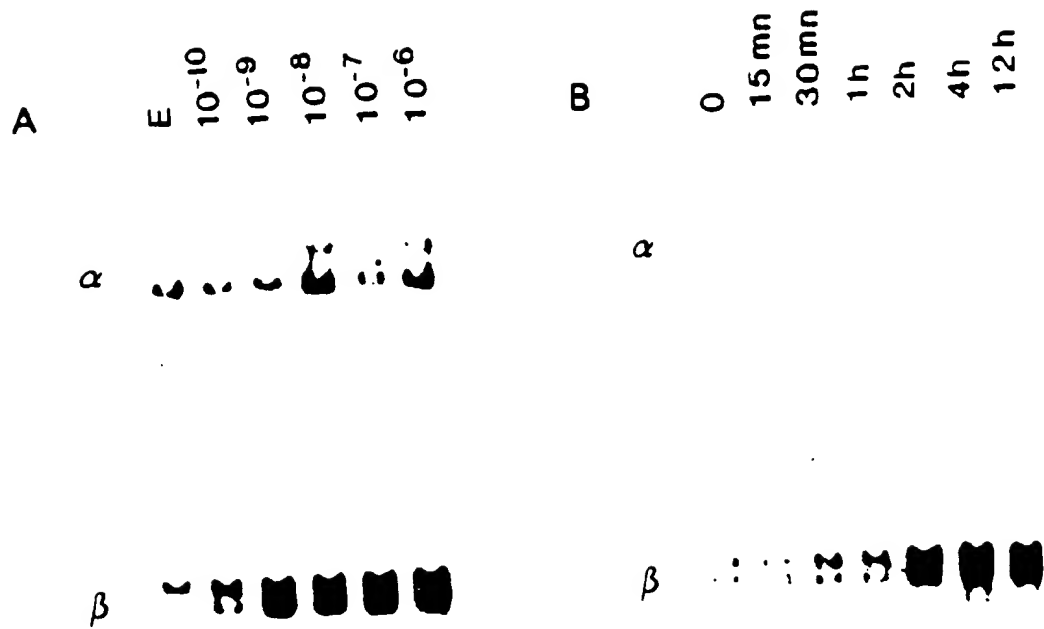


FIGURE 10



FIGURE 11

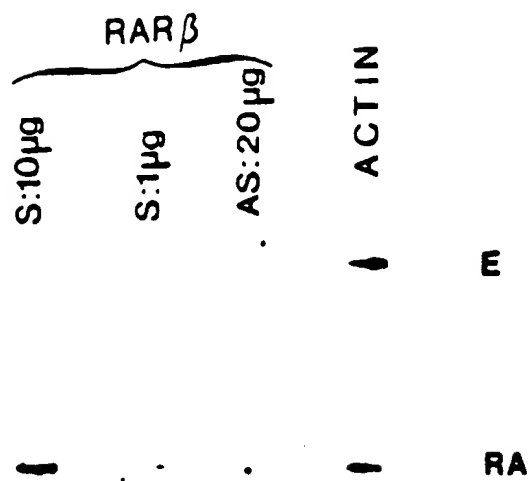


FIGURE 12

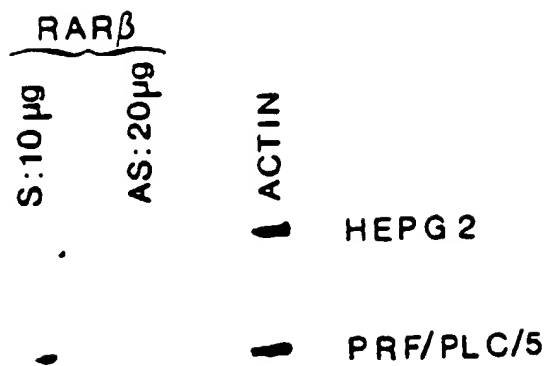


FIGURE 13

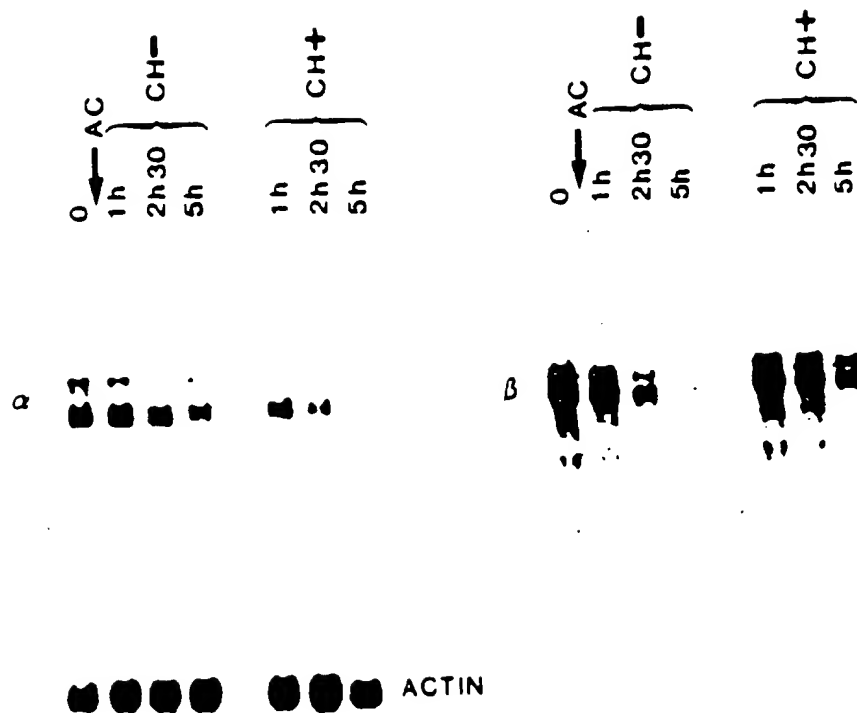


FIGURE 14

10 20 30 40 50 60 70  
 ^ ^ ^ ^ ^ ^ ^  
 CCCATGCGAGCTGTTGAGGACTGGGATGCCGAGAACCGGAGCGCATCCGAGCAGGGTCTGCTGGGCAACCGT  
 NLATII YTHIIII ALUI MNLI FOKI SFANI ACCII DPNI BSP1286  
 ALUI MNLI SFANI THAI FNUII HBOI BAN1  
 THAI NDEII SAUIIIA  
 5' end → λ13

FIGURE 15



H: Hind III, E: EcoR I, P: Pvu II, B: Bgl II, S: Sma I, Ba: BamH I.  
No Kpn I, Sac I, Sph I, Xho I are found in this DNA insert

FIGURE 16